

09-939-537.

```
Db 110 -PHORVPCAPADLLGSPVIFPPKIKDVLIMISLSPVTCVVVDSDDDVDQISFNVN 168
Qy 460 GVEHNAKTKRREQOYNSTYVSVLTVLHODMANGKVKCKNSKALPAIEKTIKSAK 519
Db 169 NVEHTATQTHREDYNSTLVSALPIQHODMMSGKEFKCKVNNRRLPSIEKTIKSKR 228
Qy 520 GQRPDPQVYTLPPRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKKTIPPVLD 579
Db 229 GPVAPAPVYVLPRAEEMTKKEFSLTCHITGLFPAELIVDNTSNKRTQNTKNTATYVLD 288
Qy 580 DGSFFLYSKLTVDKSRMQQGVFSCSVMHEALHNHYTQKSLSLSPG 625
Db 289 DGSYFMYSKLRVQKSTWERGSLFACSVYHEVTLHNHLTTKTIKSRSLG 334
```

## RESULT 43

```
501321
Ig gamma-2b chain precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 23-Jul-1999
C/Accession: S01321
R:de Waale, P.; Feys, V.; van de Voorde, A.; Moelmann, F.; Fiers, W.
Eur. J. Biochem. 176, 287-295, 1988
A/Title: Expression in non-lymphoid cells of mouse recombinant immunoglobulin directed a
A/Reference number: S01320; MUID:98329081; PMID:318116
A/Accession: S01321
A/Molecule type: mRNA
A/Residues: 1-475 <DEL>
A/Cross-references: EMBL:X13188; NID:951780; PIDN:CA31580.1; PID:951781
A/Note: this sequence was determined from the differentiated gene
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-475/Product: Ig gamma-2b chain #status predicted <MAT>
F:159-223/Domain: immunoglobulin homology <IMM>
```

```
Query March 22.8% Score 779.5; DB 2; Length 475;
Best Local Similarity 34.0%; Pred. No. 1.5e-38;
Matches 207; Conservative 66; Mismatches 159; Indels 177; Gaps 18;
```

```
Qy 30 LGKKDVELTCTAS--QKSIQPHWKNNOIKILGNQ-----SFLTKGPSKLNDRAD 81
Db 30 LARGASVYKLSCKSGTLTISYGISW-----VKQRTQGLEMIGEIYFGSNGSIFNEKFX 84
Qy 82 SRSRLW---DQGNFPLIKIKLIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOGQS 138
Db 85 GKATLTVDKSSSTAYVLMHLSLTSSEDAVYFCA---GPRQVGLLPFG-----YWGQG 132
Qy 139 LITLLESPGSSPVGQCRSPRGKNIQGGKTLVSQLELDGCTMTCTVLQNKVVERKID 198
Db 133 TLVTASAKTTPPVYPLAPFGGDTTG----- 159
Qy 199 IIVLAFOKASSIVVKEGEQVEFSPLAFVTEKLTGSGELMWQAERASSKSWITFDLN 258
Db 160 -----SSVTL---GCLVKGYPESVYLT-----WNSGSLSS----- 188
Qy 259 KEVSVKAVTQDPKIQMGKQLPLHLTLPLALPOVAGSGLTALAEKTKGLHQEVNLVVR 318
Db 189 -----VH--TFPALL-----QGGLYTMSSSVYVPS 211
Qy 319 ATQIQKULTCGVWPTSPKMLSLKLENKAKVSKREKPVVAVLPEAGMOCCLSDSGOV 378
Db 212 STWSPQVTCVAHPAS-----STTVKKLEP-----SG-- 240
Qy 379 LLESNIKILPTWSPVPCPAPRPSCKDKTHTC--PELLGSPVLPFPKPKDITMISRT 436
Db 241 -----PI-STINCP-----PCKECHCKRAPNLEGGSVIFPFPNIKDVLMISLTP 285
Qy 437 EVTCVVVDVSHEDPEVKFMYVVDGVEVNNAKTKRREQOYNSTYVSVLTVLHODMANGK 496
Db 286 KVTCAVVVDVSEDDVDQISFNVNVEVLTATQTHREDYNSTIRVSVLPIQHODMMSGK 345
```

```
Qy 497 EYCKVENKALPAIEKTIKSAKQRPDPQVYTLPPRDELTKNQVSLTCLVKGFYPSDI 556
Db 346 EFKCKVNNKDLPAIEKTIKSAKIGIYVAPQYIILSPPEQSLRSDVSLTCLAVGSPEDI 405
Qy 557 AVESWESNGQPENNYKKTIPPVLDSDGSFFLYSKLTVDKSRMQQGVFSCSVMHEALHNHYT 616
Db 406 SVEWTSNGHTEENKDTAPVLDSDGSFYIYSLKNMKTISKWEKTSFSCNVHHEGLKNYYL 465
Qy 617 OKSLSLSPG 625
Db 466 KKTISRSPG 474
```

## RESULT 44

```
506611
Ig gamma-2 chain C region (clone 32.2) - bovine (fragment)
C/Species: Bos primigenius taurus (cattle)
C/Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 16-Jul-1999
C/Accession: S06611; B31303
R:Symons, D.B.A.; Clarkon, C.A.; Beale, D.
Mol. Immunol. 26, 841-850, 1989
A/Title: Structure of bovine immunoglobulin constant region heavy chain gamma 1 and ga
A/Reference number: S06610; MUID:90097956; PMID:2513487
A/Accession: S06611
A/Molecule type: DNA
A/Residues: 1-327 <SYM>
```

```
A/Cross-references: EMBL:X16702
A/Note: the sequence was determined from the germline gene
C/Genetics:
A:Gene: Ig CH gamma 2
A:Introns: 99/1; 112/1; 219/1
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: glycoprotein; immunoglobulin; membrane protein
F:20-85/Domain: immunoglobulin homology <IMM>
F:115/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

```
Query March 22.5% Score 768; DB 2; Length 327;
Best Local Similarity 55.1%; Pred. No. 4.3e-38;
Matches 150; Conservative 40; Mismatches 64; Indels 18; Gaps 4;
```

```
Qy 371 LLSDSQGVLLSNIKV-----LPTWSTPVPCPAPRERK-----SCDKTHTCPELLG 415
Db 56 VLSGSGLYSLSWMTVPASSGQTGFCVNAHPASSIRKVXAVGVSSDCSKPNN-QHCVR 114
Qy 416 GPSVFLPFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFMYVVDGVEVNNAKTKRREQ 475
Db 115 EPSVFIPIPKPKDTLMITGPETVCVVVNVGHNDPVEQFSFVDDVETARTKPRREQ 174
Qy 476 NSTRVVSVLTVLHODMANGKVKCKNSKALPAIEKTIKSAKQRPDPQVYTLPPSRD 535
Db 175 NSTRVVSVLPIQHODMTGGEFKCKVNIKGLSIVRIISRSKGPAPRPPQVYVLDPPKE 234
Qy 536 ELTRNQVSLTCLVKGFYPSDIAVEWESNGQ--PENNYKKTIPPVLDSDGSFFLYSKLTVDK 593
Db 235 ELKSTVSLTCLVMVIGFPEVDVDMQRDQTESDKRTITPPQDLADRSYFLYSKLVYR 294
Qy 594 SRMQGNVFSCSVMHEALHNHYTQKSLSLSPG 625
Db 295 NSWQRGDTYTCVVWHEALHNHYMQSKYSKAG 326
```

## RESULT 45

```
146732
Ig gamma heavy chain constant region - rabbit (fragment)
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 21-Jan-2000
C/Accession: 146732
R:Heidmann, O.; Rougeon, F.
Nucleic Acids Res. 10, 1535-1545, 1982
A/Title: Molecular cloning of rabbit gamma heavy chain mRNA.
A/Reference number: 146732; MUID:82174328; PMID:6280149
A/Accession: 146732
A/Status: preliminary; translated from GB/EMBL/DBD
```

A:Molecule type: mRNA  
A:Residues: 1-180 <HEI>

C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
F/93-160/Domain: immunoglobulin homology <IMM>

Query Match 20.6%; Score 702; DB 2; Length 180;

Best Local Similarity 69.7%; Pred. No. 1.5e-34;

Matches 124; Conservative 26; Mismatches 28; Indels 0; Gaps 0;

QY 448 EDEVEKNNVVDGVEVHNAKTKPREQYNSTYRVSVLTIVLHODWLNKGEYKCKVSNKAL 507

DB 2 DDEVOFTYINNEQRTARPLREQQFNSTIRVSTLPIAHQDWLRGKEFKCKVANKAL 61

QY 508 PAIEKTIKAKQGPPEPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPE 567

DB 62 PAIEKTIKAKQGPPEPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPE 121

QY 568 NNKKTTPVLDSDGSFPLYSKLTVDKSRMOQGNVFSQVHMEALHNHYTQKSLSLSPG 625

DB 122 DNKKTTPVLDSDGSFPLYSKLTVDKSRMOQGNVFSQVHMEALHNHYTQKSLSLSPG 179

#### RESULT 46

A36040  
Ig heavy chain V-III region (ART) - human (fragments)

C/Species: Homo sapiens (man)  
C/Date: 16-Nov-1990 #sequence\_revision 16-Nov-1990 #text\_change 21-Jan-2000

C/Accession: A36040  
R/Bullitz, M.; Weiss, D.T.; Solomon, A.

Proc. Natl. Acad. Sci. U.S.A. 87, 6542-6546, 1990

A/Title: Immunoglobulin heavy-chain-associated amyloidosis.

A/Reference number: A36040; MUID:90370821; PMID:2118650

A/Accession: A36040  
A:Status: preliminary

A:Molecule type: protein  
A:Residues: 1-218 <EUL>

C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin

F/132-199/Domain: immunoglobulin homology <IMM>

Query Match 16.8%; Score 574.5; DB 2; Length 218;

Best Local Similarity 77.5%; Pred. No. 6.2e-27;

Matches 110; Conservative 3; Mismatches 12; Indels 17; Gaps 1;

QY 484 VLTVLHODWLNKGEYKCKVSNKALPAIEKTIKAKQGPPEPOVYTLPPSRDELTKNQVS 543

DB 94 VAAALLTDMQGG-----TLVTVSSSGQPREPOVYTLPPSRDELTKNQVS 136

QY 544 LTCLVGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVFS 603

DB 137 LTCLVGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVFS 196

QY 604 CSVMHEALHNHYTQKSLSLSPG 625

DB 197 CSVMHEALHNHYTQKSLSLSPG 218

RESULT 47

S69340  
Ig heavy chain VHIII-D-JH-CH3 region - human

C/Species: Homo sapiens (man)  
C/Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 21-Jan-2000

C/Accession: S69340  
R/Khamlich, A.A.; Aucouturier, P.; Freund, H.; Cogné, M.

Eur. J. Biochem. 229, 54-60, 1995

A/Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.

A/Reference number: S69339; MUID:95262687; PMID:7744049

A/Accession: S69340  
A:Status: preliminary

A:Molecule type: mRNA  
A:Residues: 1-249 <KHA>

C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
F/162-229/Domain: immunoglobulin homology <IMM>

Query Match 16.8%; Score 572.5; DB 2; Length 249;

Best Local Similarity 62.9%; Pred. No. 9.7e-27;

Matches 117; Conservative 9; Mismatches 23; Indels 37; Gaps 3;

QY 460 GVEVHNAKTKPREQYNSTYRVSVLT-----VLHODWLNKGEYKCKVSN 504

DB 80 GDSVGRFTISRDNSENQILDMNSLTADTAVYCAKGVSYIH----- 124

QY 505 KALPAIE-----KTIKAKQGPPEPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE 559

DB 125 --LPSAIHFWGQGVTVYVSSGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE 182

QY 560 WESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVFSQVHMEALHNHYTQKS 619

DB 183 WESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVFSQVHMEALHNHYTQKS 242

QY 620 LSLSPG 625

DB 243 LSLSPG 248

#### RESULT 48

S14236  
Ig gamma-1 chain C region (15C5) - mouse (fragment)

C/Species: Mus musculus (house mouse)  
C/Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C/Accession: S14236  
R/Vandamme, A.M.; Bullens, F.; Bernar, H.; Nelles, L.; Lijnen, R.H.; Collen, D.

Eur. J. Biochem. 192, 767-775, 1990

A/Title: Construction and characterization of a recombinant murine monoclonal antibody

A/Reference number: S14236; MUID:91006173; PMID:2209622

A/Accession: S14236  
A:Molecule type: mRNA

A:Residues: 1-152 <VANA>

C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C/Keywords: immunoglobulin

F/65-132/Domain: immunoglobulin homology <IMM>

Query Match 16.6%; Score 566; DB 2; Length 152;

Best Local Similarity 61.6%; Pred. No. 1.2e-26;

Matches 93; Conservative 33; Mismatches 25; Indels 0; Gaps 0;

QY 475 YNSTYRVSVLTIVLHODWLNKGEYKCKVSNKALPAIEKTIKAKQGPPEPOVYTLPPSR 534

DB 1 FNSTFRSVSELPIMHODWLNKGEYKCKVSNKALPAIEKTIKAKQGPPEPOVYTLPPSR 60

QY 535 DELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKS 594

DB 61 EQMAKDKVSLTCLVTFPEPDITVEWQNGQPENNYKTPPVLDSDGSFPLYSKLTVDKS 120

QY 595 RMQGNVFSQVHMEALHNHYTQKSLSLSPG 625

DB 121 NMEAGNTFTCSVHBEGLHNHYTQKSLSLSPG 151

RESULT 49

S04845  
Ig heavy chain precursor - African clawed frog (fragment)

C/Species: Xenopus laevis (African clawed frog)  
C/Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 26-Aug-1999

C/Accession: S04845; S05695  
R/Amemiya, C.T.; Haire, R.N.; Litman, G.W.

Nucleic Acids Res. 17, 5388, 1989

A/Title: Nucleotide sequence of a cDNA encoding a third distinct Xenopus immunoglobulin

A/Reference number: S04845; MUID:89345103; PMID:2503814

A/Accession: S04845  
A:Molecule type: mRNA

A:Residues: 1-549 <AME>



```

Db      32 KPGSSVKSCKAGGTFSSAISWQAQGGLEMM-GGIIPIFGT-ANYAQKFGQRTVI 89
      79 RADSRSLMDQGNFPIIKKLTIEDSDTYICEVEDQKEEVQL---VFGLTANSDFHL-- 133
      90 TADESTS-----TAYMELSLRSEDYAVVYC-----AKTGILPYSSGYPMNDYYVYG 138
Qy      134 --LOGGSLTLTLESPGSSPS-----VQCR----- 156
      139 MDVWGCGTTVTVSSGSASAFITLPPVYSCENSPEDTSVAAGLADPFLPDSTIFSMKTKN 198
Db      157 -----SPRG--KNIQGGKTLVSQLELQD-----SGT---WTCTVLQNKQKVEFKIDIVL 202
      199 NSDISSTRGFPVLRGKVAATSQVLLPSKDVMQGDENHVCVQHPNGKNGKAVPLPIY 258
Qy      203 AF--QKASSIYVKKKEQVQFSPFLATYKLTGSGELMWQARASSK---SWITPDLK 257
      259 AELPPKVSVPVPRDG---FFGNP-----RSKSKLICQAAGFSPROIQVSW----- 301
Qy      258 NKEVSVRVTPQDKLQMGKTLPLHLTPQALPQVAGSGLTLALAKTKGLHQEVNLVVM 317
      302 -----LREGQV-----GSGVTTDOVQMAKES----- 324
Qy      318 RATOLQKLTCEVWGPTSPKMLSLKLENKAVSKREKRVNVLNPEAGMOCILSDSQ 377
      325 -----GPTTKVLTSTLTIKESD-----WL---SQSMFTCRVDHRL 357
Db      378 VLE-SNIKVLPTWSPVPCPADEPKSCDKHTCPPELLGSPVFLPPPKKDTLMISRP 436
      358 TFOQNASMCMCPDODTAI-----RVFALPPS-FASIFLTKST 393
Qy      437 EVTCVAVDVSHEDPEVFNMYVDGVEVHNAKTPREEQNSTYRVSVLTVLHODMLNGK 496
      394 KTLCLVTLDTLTYD-STVISTRONGEAVKTHNTISEHPRATTSANGEASICEDDMNSGE 452
Db      497 EYCKVSNKALPAPIEKTISKAGQP-REPQVYTLPPSRDELTKQVSLTCLVKGYPS 554
      453 RFTCTVHTDLPSPKQTSIRPKGVALHRPDVLLPAPRQMLRESATITCLVTVGSPA 512
Qy      555 DIAVESNSQGP--ENNYKTTTPVLD--SDGSFFLYSKLTVDKSRMQGVNFGSCVWHEA 610
      513 DVFQWQWQROPPLSPPEKYVASPMPEQAPGRYFAHSILTIVSEEWNTGETYTCVVAHEA 572
Db      611 LHNHYTOKSLSLSPGLQDETCAEADQDELGLMTT 646
      573 LPRKVTERTVDKS-----TEGEVSADDEGEFENLMAT 603

```

## RESULT 52

S38864

Ig epsilon chain C region - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 24-May-2001

C/Accession: S38864

R/KIPC, B.; Becker, W.; Schlaak, M.

Submitted to the EMBL Data Library, November 1993

A:Description: Combination of a defined specificity and desired isotype by cloning of ar

A:Reference number: S38864

A/Accession: S38864

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-546 &lt;RIP&gt;

A/Cross-references: EMBL:Z27397; NID:g416537; PIDN:CAAB1788.1; PID:g940782

C:Superfamily: immunoglobulin C region; immunoglobulin homology

F:353-421/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 11.0%; Score 374; DB 2; Length 548;

Best Local Similarity 23.3%; Pred. No. 1,2e-14;

Matches 155; Conservative 105; Mismatches 206; Indels 198; Gaps 31;

```

Qy      30 LGKGGDTVELTCTAS--OKKSIOFHW-----KNSNOIKILGNGSFLTKGSPSKLN--- 77
      11 LKVGSGSLKLSAASGLTFSSYGSWVRQIIPDKRLLEWAVATISSGTY-TYPPDSVGRFT 69

```

```

Qy      78 -RADSRSLMDQGNFPIIKKLTIEDSDTYICEVEDQKEEVQLVFGLTANSDFHLQ 136
      70 IERDAKNTLY-----LQMSLSKSEDTAMYCAAGVSTMRFAWM-----G 111
Qy      137 QGLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG-----TWCTVQLQ 188
      112 OGTLTVASAGKTTTPSVYPLAG-----SAAQNSMWTLCGLVGVYPPETVLMNSGSL- 166
Qy      189 NQKVEFKIDIVVLAQKASSIYVKKKEQVQFSPFLATYKLTGSGEL---WQME 243
      167 -----SSGVH-----TPPAVLQSLDYTLISSVTPSSTWPEE 198
Qy      244 RASSKSWITPDLKXKEVSVRVTPQDKLQMGKTLPLHLTPQALPQVAGSGLTLALBA 303
      199 TVTCN---VAHASTKVDKIVPRD---CGCK-----PCIYV----- 230
Qy      304 KTGKLTQHEVNLV-----VMRATQOLKLTCEVWGPTSPKMLSLKLENK----- 348
      231 -----EVSVFIFPPPKQDVLTRSTIQ--LYCFYGHILNDVSVSMIMDDREITDTLA 281
Qy      349 -----AKVSKREKRVNVLNPEAGM-----CLSDSQVLESNIKVLPTWSTVPCP 397
      282 QTVLKEBQKLASTCSKLNITEQWMSSESTFCVTSQGVLYLAHTR-----CP 331
Qy      398 ADEPKSCDKHTCPPELLGSPVFLPPPKKDTLMISRTPEVTCVVVDV-SHEDPEVKFN- 455
      332 DHEPR-----GVITYLIPSPLD--LYQNGAKRLCTLVLDLSEKKNVNTVMQ 377
Qy      456 -----WYVDGVEVHNAKTPREEQNSTYRVSVLTVLHODMLNGKEXKCVSNK 505
      378 EKTISVASAGWY---TKHN-----NATTSITSLIPVAKWIMIGGYQCIVDHP 424
Db      506 ALPAPIEKTISKAGQPREQVYTLPPSRDELTKQVSLTCLVKGYPPDIAVEMESNQ 565
      425 DEPKPIVRITTPGQSAPEVYVPPPEE--SEDKRTITCLIQNPPEDISVQWGDGK 483
Qy      566 PENN--YKTTTPVLDSDGS---FFLYSKLTVDKSRMQGVNFGSCVWHEALHN-HYTOKS 619
      484 LISNSGHSITTP-LKNSGNSRGEFTSRLEVAKTLMTOKQKQFVQVYHHLQKRLKLT 542
Qy      620 LSLS 623
      543 ISTS 546

```

## RESULT 53

EHMS

Ig epsilon chain C region (version 1) - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 16-Jul-1999

C/Accession: A02144

R/LIU, F.T.; Albrandt, K.; Sutcliffe, J.G.; Katz, D.H.

Proc. Natl. Acad. Sci. U.S.A. 79, 7852-7856, 1982

A:Title: Cloning and nucleotide sequence of mouse immunoglobulin epsilon chain cDNA.

A:Reference number: A02144; M0ID:83117774; PMID:6818553

A/Accession: A02144

A/Molecule type: mRNA

A/Residues: 1-388 &lt;LIU&gt;

A/Cross-references: GB:000476; NID:g194875; PIDN:AAA38085.1; PID:g387220

C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (ka

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin

F:1-44/Domain: immunoglobulin homology (fragment) &lt;IML&gt;

F:81-149/Domain: immunoglobulin homology &lt;IM2&gt;

F:186-254/Domain: immunoglobulin homology &lt;IM3&gt;

F:290-361/Domain: immunoglobulin homology &lt;IM4&gt;

F:10,51,62,133,205,228,332,382/Binding site: carbohydrate (asn) (covalent) #status pred

Query Match 10.8%; Score 367.5; DB 1; Length 388;

Best Local Similarity 26.7%; Pred. No. 1,9e-14;

Matches 118; Conservative 73; Mismatches 138; Indels 113; Gaps 20;



```

Oy 222 SFLPAFTVEKLTGSGELMWQAEERASSKSWITFDLKNKEVSKRVLTDDPKLQMGKLT--- 278
Db 15 NFPALGSELKVTTSQVTSW-----GKSANQFC-----HYTHPPSNESSTILVR 59
Oy 279 PLHLTLPALPOYASGNLTALAEKTKLHOENVLVVRATLOLOKLTCEVWGPSTPKL 338
Db 60 PNVITEP-----LTLEL-----LHSSCPNPFHST---IQLYCFIYGHILNDV 98
Oy 339 MLSLKEKE-----AKYSKREKPVWVNLPEAGMWQ-----CLISDGOVLLE 381
Db 99 SVSNLMMDREITDITLAQVTLKEBSKLASTGSKLITTEQOMWSESTFCCKVTSQGVLYLA 158
Oy 382 SNIKVLPWSTPVPFCAPAPKSCDKTHTCPPELLGSPVFLPPPKKDTLMSRTPPEVTCV 441
Db 159 HTRR-----CPDHEPR-----GVITYLLPSPPLD-LYQNGAPKLTCL 194
Oy 442 VVDV-SHEDPEVKR-----WYVDGEVHNATKPREQYNSTYRVSVLYVLA 489
Db 195 VVDESEKGVVVTWNOEKRTSVSASQWY---TKHN-----NATTSITSLPVYA 241
Oy 490 QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVK 549
Db 242 KQWLEGGYQCIVDPDPFKPIVSTKTPGQSAPEVYVPPEER-SEDKRTLTCLIQ 300
Oy 550 GFVPSDIAVEWESNQGPENN--YKTPPVLDSDGS---FPLYSKLTVDKSRMQQGNVFC 604
Db 301 NFPEDISVQWLGSGKLSNSGHSHTTP-LKSNNGSNQGFIFSLRLVAKTILMTQKQPTC 359
Oy 605 SVMGEALHN-HYTKQSLSLSPG 625
Db 360 QVHEHALQPKRKLEKTIISTSLG 381

```

## RESULT 54

```

Ehrt
Ig epsilon chain C region - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Dec-1992 #sequence revision 17-Dec-1992 #text_change 16-Jul-1999
C:Accession: A93442; A90937; X02143
R:Helman, L.; Peterson, U.; Engstrom, A.; Karlsson, T.; Bennich, H.
Nucleic Acids Res. 10, 6041-6049, 1982
A:Title: Structure and evolution of the heavy chain from rat immunoglobulin E.
A:Reference number: A93442; MUID:83064537; PMID:6292865
A:Accession: A93442
A:Molecule type: mRNA
A:Residues: 1-429 <HEL>
A:Experimental source: strain LOU/C/Wel, immunocytoma IR2
R:Kindvogel, W.R.; Reddy, E.P.; Moore, J.M.; Faust Jr., C.H.
DNA 1, 335-343, 1982
A:Title: A cloned cDNA probe for rat immunoglobulin epsilon heavy chain: construction,
A:Reference number: A90937; MUID:83182019; PMID:6820340
A:Contents: myeloma IR162
A:Accession: A90937
A:Molecule type: mRNA
A:Residues: 'N', 169-307, 'L', 309-342 <KIN>
C:Complex: An immunoglobulin heterotrimeric subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 12
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
F:19-80/Domain: immunoglobulin homology <IM1>
F:118-186/Domain: immunoglobulin homology <IM2>
F:223-291/Domain: immunoglobulin homology <IM3>
F:327-398/Domain: immunoglobulin homology <IM4>
F:46,99,170,240,265,369,419/Binding site: carbohydrate (Aan) (covalent) #status predicted

```

```

Query Match 10.6%; Score 363.5; DB 1; Length 429;
Best Local Similarity 30.6%; Pred. No. 3.8e-14;
Matches 91; Conservative 60; Mismatches 99; Indels 47; Gaps 12;
Oy 342 LKLENKAKVSKREKPVWNLPEAGMWQ-----CLISDGOVLLESNIKVLPWSTPVP 395
Db 156 IKEKGKLAIVSYR-----LNIIDQOMWSESTFCCKVTSQGE-----NYMAHTRR 199

```

```

Oy 396 CPAPKSCDKTHTCPPELLGSPVFLPPPKKDTLMSRTPPEVTCVVDVSHEDPEVKFN 455
Db 200 CSDEPR-----GVITYLLPSPPLD-LYENGKPKLCIVLDESE-NITVT 244
Oy 456 WYVDGEVHNATKPREQYNSTYRVSVLTVLHQMVLNGKEYKCKVSNKALPAPIEKT 515
Db 245 WVRERKKSIGSASQSRKHNATTSITSLIPVDADWIEGGYQCRVDHPFPKPIVRSI 304
Oy 516 SKAGQPREPOVY-LPPSRDELTKNQVSLTCLVKGFVPSDIAVEW--ESNGQPENNYKT 572
Db 305 TKAGKASAPEVYVFLPPEEBE--KDKRTLTCLIQNFPEDISVQWLODSGLIFKSGHST 362
Oy 573 TPVLDSDGS---FPLYSKLTVDKSRMQQGNVFCVMEHALHN-HYTKQSLSLSPG 625
Db 363 TTP-LKXNGSNQGRFISRLLEVTLMTQKQFCRVIHEALRPRKLERIISTSLG 418

```

## RESULT 55

```

Ig mu chain - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 23-Jul-1999
C:Accession: S25705
R:Patil, S.; Nau, F.
Mol. Immunol. 29, 829-836, 1992
A:Title: Isolation and sequence of a cDNA coding for the immunoglobulin mu chain of the
A:Reference number: S25705; MUID:92342148; PMID:1655560
A:Accession: S25705
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-592 <PART>
A:Cross-references: EMBL:X59994; NID:G1269; PIDD:CAA42611.1; PIDD:G1270
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:483-554/Domain: immunoglobulin homology <IM1>

```

```

Query Match 10.6%; Score 363.5; DB 2; Length 592;
Best Local Similarity 22.6%; Pred. No. 5.7e-14;
Matches 155; Conservative 97; Mismatches 233; Indels 181; Gaps 29;
Oy 6 PFRHLVLVQL---ALLPAATQGNKVVLAGKGDVETCTASQKSIQFH--WKNNOIK 60
Db 3 PLMTLLFVLSAPRGVLSQVQLQESGPELVKPSFTLSLCTVSGSGLTVNHNWIRQASGK 62
Oy 61 ILNGQSFLLTKGPKLNDRADSRSL--WDQGNFPLIKLTKEDSDTYICEVEDQKEE 117
Db 63 MPWLVGVEKGGNTYVNPALKSRISIRDTYSKQSVLSLSMIDIDRAVYYC----- 114
Oy 118 VQLLVFLGLTANSDDTHLQ-----GQS/LTLF---ESPPGSSPSVOC-RSPRGKNIQ-- 164
Db 115 -----ARSAGVFLADVDIWKGILLVTVSSESESHPRVYFPLVSCVSSPSDENTVALG 166
Oy 165 -----GGKTLVSQQL---ELQDSGTWCT--VLQNKQVVEFKIDIV 200
Db 167 CLARDPVPNVSPSKKNNSTVSSERFWTPPEVLRDGLWASSQVLAHSSSTFGGTGYL 226
Oy 201 VLAFO--KASSIYKKEGEOVERSPFLAFT---VEKLTSGELIMQAEASSSKSWITF 254
Db 227 VCEQVHQHKGEDVGHKGPPREVEVLSPVSVFVPCNSLSGNG-----NSKSLICQAT 279
Oy 255 DLKMKVEVY---KRVTDPPKLQMGKULPLHLTLPOALPOYAGSGLTLALEAKTGKL 308
Db 280 DFSEKQISLSWFRDGRKIVSD-----ISBQVETVOSSPTTYR 317
Oy 309 HOEVNLVWRATLOKNTLCEVWGPTSPKLMLSLKENKEAKVSKREKPVWNLPEAGMW 368
Db 318 AVSVLTITREEMLSQSAVYTCQV-----ENKKT----- 345
Oy 369 QCLISDGOVLLESNIKVLPWSTPVPFCAPAPKSCDKTHTCPPELLGSPVFLPPPKPK 428
Db 346 -----FGKNAS-----SSCDATPPSP--IG--VFTIPPSFAD 373
Oy 429 TLMISRTPEVTCVVDVSHEDPEVKFN--VVDG--VEVHNATKPREQYNSTYRVSVL 485

```



A: Molecule type: DNA  
 A: Residues: 436-476 <EAR>  
 A: Cross-references: GB:V00816; GB:J00444; NID:G52343; PIDN:CAA24197.1; PID:G52344  
 R: Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bord, M.; Hood, L.; Wall, R.  
 Cell 20, 303-312, 1980  
 A: Title: Two mRNAs with different 3' ends encode membrane-bound and secreted forms of Ia  
 A: Reference number: A37517; MUID:8022873; PMID:6771012  
 A: Contents: MOPC 104B  
 A: Accession: A37517  
 A: Molecule type: mRNA  
 A: Residues: 410-476 <ROG>  
 A: Cross-references: GB:V00821; NID:G52355; PIDN:CAA24202.1; PID:G817972  
 R: Kawakami, T.; Takahashi, N.; Honjo, T.  
 Nucleic Acids Res. 8, 3933-3945, 1980  
 A: Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison with  
 A: Reference number: A02166; MUID:81076590; PMID:6255422  
 A: Accession: B02166  
 A: Molecule type: DNA  
 A: Residues: 1-435, 'GKPTLVNLSIMSDGTCY' <KAN>  
 A: Comment: The sequence of residues 1-409 was assumed to be identical with the counterpart  
 C: Comment: During differentiation, B lymphocytes switch from expression of membrane-bound  
 B: C: Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap  
 hain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1a  
 C: Superfamily: immunoglobulin C region; immunoglobulin homology  
 C: Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob  
 F: 21-91/Domain: immunoglobulin homology <IMM1>  
 F: 129-201/Domain: immunoglobulin homology <IMM2>  
 F: 239-307/Domain: immunoglobulin homology <IMM3>  
 F: 346-417/Domain: immunoglobulin homology <IMM4>  
 F: 436-476/Domain: carboxyl-terminal <CTS>  
 F: 14/Disulfide bonds: interchain (to light chain) #status predicted  
 F: 28-89, 136-199, 246-305, 353-415/Disulfide bonds: #status predicted  
 F: 46-211, 243, 258, 281/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F: 216/Disulfide bonds: interchain (to heavy chain) #status predicted  
 F: 293/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.5%; Score 359.5; DB 1; Length 476;  
 Best Local Similarity 22.7%; Pred. No. 7.4e-14;  
 Matches 127; Conservative 80; Mismatches 187; Indels 165; Gaps 19;

Dy 144 ESPGSSPVQCRSP-RGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEKIDIVL 202  
 3 QSPFNVPFLVCEPSLDKLNVAAGCLARDFLPSTISFTN----- 43  
 Dy 203 AFQRASSIVYKKEGEQVEFPFLAFYVEKLTGSGELMWQAEKASSSKSWTFDLKNEVS 262  
 44 -----YQNTTEVIQ-----GIRTFPLRTGCKYLA 68  
 Dy 263 VKRVTQDPKLOM-----GKKLPLHLTLPLQALPQVAGSGLTLALAKTKGL 308  
 69 TSQVLBPKSLIEGSDLVCKIKHYGKRNRLHPITP-----AVAEW 110  
 Dy 309 HQEVNLVV-----MRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWLN 362  
 111 NPNNVNFVPPRDPGSGAPRKSXKLIICENTNTPKPIYWSMKDKLVESGFTTDPVILEN 170  
 Dy 363 -----PEA-----GMQCLISDSGQVLESNIVLPTWSTVPAPAP 401  
 171 KGSPTQRYKIVISTLTISEIDWLNINLVYTCRVHRGLFTLK----- 210  
 Dy 402 KSCDKTHTCPRELLGSPS---VFLPRPKDITLMISTPRTVCVNVDSHEDPEVKNWY 457  
 211 ---NVSTC---AASPSTDIITFTTPSPFAD-IFLSKSANITCLVSNLATE-TLNLSWA 262  
 Dy 458 VDGVEVNAKTPREEQYNSTYRVSVLTVLHODWLNGKEKCKVSKALPAPIEKTISK 517  
 263 SQSEPLEETKIKIMESHNGTFSAKGVASVCEBNNRKEVCVYTHRDLPSPQKRTISK 322  
 Dy 518 AKGQPRE-----POVYTLPSRDEL-T-KNOVSLTCLVKGFPYSDIAVESNNGQ--DENN 569  
 323 -----FNEVHKHPRVAVYLLPAREQDNLNRESATVTCVGVGSPADISQVMIQRLGQLPQEK 378

Dy 570 YKTPPVLD--SDGSFFLYSKLTVDKSRMOQGNVFSQVMEALHNHYTKSLSPQLQ 627  
 379 YVTSAPNPEPAPPEFYFTSLTTEEBMNSGERTYCVVGHFALPHVTERTVKSKS---- 434  
 Dy 628 LDFTCAPADGDELGLMTT 646  
 435 -TEGEVNAEPEGFENLMTT 452

RESULT 59  
 S15590  
 Ig heavy chain - human  
 C: Species: Homo sapiens (man)  
 C: Date: 18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
 A: Accession: S15590  
 R: Neale, G.A.M.; Kitchingman, G.R.  
 Nucleic Acids Res. 19, 2427-2433, 1991  
 A: Title: mRNA transcripts initiating within the human immunoglobulin mu heavy chain enh  
 A: Reference number: S15590; MUID:9125286; PMID:1904154  
 A: Accession: S15590  
 A: Status: preliminary  
 A: Molecule type: mRNA  
 A: Residues: 1-474 <NEA>  
 A: Cross-references: EMBL:X58529  
 A: Note: the authors translated the codon CAA for residue 265 as Glu  
 C: Superfamily: immunoglobulin C region; immunoglobulin homology  
 C: Keywords: heterotetramer; immunoglobulin  
 F: 344-415/Domain: immunoglobulin homology <IMM4>

Query Match 10.5%; Score 359; DB 2; Length 474;  
 Best Local Similarity 26.7%; Pred. No. 7.9e-14;  
 Matches 86; Conservative 65; Mismatches 121; Indels 50; Gaps 11;

Dy 332 GPISPKMLSLKENKAKVSKREKPVVNLNPEAKMQCLISDSGQVLE-SNIKVLPTW 390  
 172 GPITYKVTSLTIKESD-----WL---SQSMFTCRVHRGLTFQVNASMCVPDQ 218  
 Dy 391 STVPKCAPRKSCDKTHTCEBLLGSPSVLFPKPKDITLMISTPRTVCVNVDSHEDP 450  
 219 DTAI-----RVFAIPPS-FASIFLTSTKLTCLVTLDTTYD- 253  
 Dy 451 EVKENWYVDGVEVNAKTPREEQYNSTYRVSVLTVLHODWLNGKEKCKVSKALPAP 510  
 254 SVTISWRQNGQAVKHTNISESHNATPFAVAGASICEDDMNSGERTYCVYTHRDLPSP 313  
 Dy 511 IETKISAKQOP-REPQVYTLPSRDEL-T-KNOVSLTCLVKGFPYSDIAVESNNGQ-- 566  
 314 LKQITSRPKGVALHRPQVLLPAREQDNLNRESATITCLVTGSPADVPVQMQRGPLS 373  
 Dy 567 ENNYKTPPVLD--SDGSFFLYSKLTVDKSRMOQGNVFSQVMEALHNHYTKSLSP 624  
 374 PEKYVTSAPNPEPAPPEFYFTSLTTEEBMNSGERTYCVVGHFALPHVTERTVKSKS- 432  
 Dy 625 GLDFTCAPADGDELGLMTT 646  
 433 -----TEGEVNAEPEGFENLMTT 450

RESULT 60  
 S37768  
 Ig mu chain C region - human  
 C: Species: Homo sapiens (man)  
 C: Date: 12-Feb-1998 #sequence\_revision 12-Feb-1998 #text\_change 23-Jul-1999  
 A: Accession: S37768  
 R: Haridraath, N.; Donadel, G.; Sigoune, G.; Nockins, A.L.  
 Mol. Immunol. 30, 111-112, 1993  
 A: Title: Comparison of complete nucleotide sequence of the human IgM heavy chain const  
 A: Reference number: S37767; MUID:93109368; PMID:8417370  
 A: Accession: S37768  
 A: Molecule type: mRNA  
 A: Residues: 1-453 <HAR>  
 A: Cross-references: EMBL:X67301; NID:G38407; PIDN:CAA47714.1; PID:G38408  
 A: Experimental source: cell line Ab 63

C, Genetics:  
 A: Map position: 14q32  
 C: Superfamily: immunoglobulin C region; immunoglobulin homology  
 C: Keywords: glycoprotein; heterodimer; immunoglobulin  
 F: 21-90/Domain: immunoglobulin homology <IMM1>  
 F: 127-199/Domain: immunoglobulin homology <IMM2>  
 F: 237-305/Domain: immunoglobulin homology <IMM3>  
 F: 344-415/Domain: immunoglobulin homology <IMM4>  
 F: 14/Dsulfide bonds: interchain (to light chain) #status predicted  
 F: 28-88, 134-197, 244-303, 351-413/Dsulfide bonds: #status predicted  
 F: 16, 209, 272, 279, 440/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F: 214, 452/Dsulfide bonds: interchain (to heavy chain) #status predicted  
 F: 291/Dsulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.5%; Score 357.5; DB 2; Length 453;  
 Best Local Similarity 27.6%; Pred. No. 9.1e-14;  
 Matches 83; Conservative 62; Mismatches 11; Indels 45; Gaps 10;

332 GPTSPKLMSTIKENKAKVSKREKPVWVNPAGMOCILSDSGQVLE-SNIXVLPW 390  
 172 GPTTVVTSTLTKESD-----WL---SQSMFTCKVDHRLGTFQOMSSMCVPO 218  
 391 STEVPCAPAEPSKCDKTHTCPELLGSPVFLPPPKPDLTMSRTPEVTCVVDVSHEDP 450  
 219 DPAI-----RVPALPPS-FASIFLTKSKTKLTCLVTLDTYD- 253  
 451 EVKFNMYVDGVHNAKTKPREQVNSTRVSVLTVLHODMLNKEKCKVSNKALPAP 510  
 254 SVTISWTRONGEAVKTHNTISEHPNATFSAVEASICEBDNMSGERFTCTVHTDLPSP 313  
 511 IETTKAKAGQP-REPOVYTLPPSRDELTKNOVSLTCLVKGVPYSDIAYEMESNGQP- 566  
 314 LKQTFISPKVLAHRPVDVLLPAREQMLNRESATITCLTGTSPADVOMORQDPDS 373  
 567 ENNYKTPPVLD--SDGSFPLYSKLTIVDKSRWQGNVFCSCVHGAHLNHYTKSLSP 624  
 374 PEKYVTSAPPEPQAGRYFAHSILTVSEEMTGTGYTCVVAHEALPKRVTERTYDKST 433

QY 625 G 625  
 DB 434 G 434

RESULT 61  
 160082  
 CD4 receptor - human (fragment)  
 C: Species: Homo sapiens (man)  
 C: Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 23-Jul-1999  
 C: Accession: 160082  
 R: Zverev, V.V.; Sidoren, A.V.; Nedospasov, S.A.; Malushova, V.V.; Udalova, I.A.; Andzha  
 Vopr. Virusol. 40, 100-102, 1995  
 A: Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene].  
 A: Reference number: 160082; MUID: 95407135; PMID: 7676667  
 A: Accession: 160082  
 A: Status: preliminary; translated from GB/EMBL/DBD  
 A: Molecule type: mRNA  
 A: Residues: 1-71 <RES>  
 A: Cross-references: GB: S79267; NID: g1086922; PIDN: AAB35273.1; PID: g1086923  
 C: Genetics:  
 A: Introns: 17/1  
 C: Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

Query Match 10.5%; Score 357; DB 2; Length 71;  
 Best Local Similarity 98.6%; Pred. No. 9.2e-15;  
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLTQALTPAATGKNKVTGKGGDTVELCTASOKKSIOFHKNKSNQIK 60  
 DB 1 MNRGVPFRHLVLTQALTPAATGKNKVTGKGGDTVELCTASOKKSIOFHKNKSNQIK 60

QY 61 ILNGSGFLTK 71  
 DB 61 ILNGSGFLTK 71

RESULT 62  
 EHRU  
 Ig epsilon chain C region - human  
 C: Species: Homo sapiens (man)  
 C: Date: 31-Mar-1981 #sequence\_revision 13-Jun-1983 #text\_change 16-Jul-1999  
 C: Accession: A22771; A23195; PH1214; A93491; A90824; A94418; B93933; S02438; A53116; C4  
 R: Flanagan, J.G.; Rabbitt, T.H.  
 EMO J. 1, 655-660, 1982  
 A: Title: The sequence of a human immunoglobulin epsilon heavy chain constant region gene  
 A: Reference number: A22771; MUID: 84236029; PMID: 6234164  
 A: Accession: A22771  
 A: Molecule type: DNA  
 A: Residues: 1-428 <FLA>  
 A: Cross-references: GB: L00022; GB: J00227; GB: V00555; NID: g185035  
 R: Ueda, S.; Nakai, S.; Nishida, Y.; Hisejima, H.; Honjo, T.  
 EMO J. 1, 1539-1544, 1982  
 A: Title: Long terminal repeat-like element flank a human immunoglobulin epsilon pseudo  
 A: Reference number: A23195; MUID: 84207910; PMID: 6327276  
 A: Accession: A23195  
 A: Molecule type: DNA  
 A: Residues: 2-428 <UED>  
 A: Cross-references: GB: J00222; NID: g184755  
 R: Zhang, K.; Saxon, A.; Max, E.E.  
 J. Exp. Med. 176, 233-243, 1992  
 A: Title: Two unusual forms of human immunoglobulin E encoded by alternative RNA splicing  
 A: Reference number: PH1214; MUID: 92308839; PMID: 1613458  
 A: Accession: PH1214  
 A: Molecule type: DNA  
 A: Residues: 320-428 <ZHA>  
 A: Cross-references: EMBL: X63693; GB: S18668; NID: g32987  
 R: Seno, M.; Kurokawa, T.; Ono, Y.; Ono, H.; Sasada, R.; Igatahi, K.; Kikuchi, M.; Sugi  
 Nucleic Acid Res. 11, 719-726, 1983  
 A: Title: Molecular cloning and nucleotide sequencing of human immunoglobulin epsilon ch  
 A: Reference number: A93491; MUID: 83168897; PMID: 6300763  
 A: Accession: A93491  
 A: Molecule type: mRNA  
 A: Residues: 1-428 <SEN>  
 A: Cross-references: GB: L00022; GB: J00227; GB: V00555; NID: g185035  
 R: Max, E.E.; Batey, J.; Ney, R.; Kirsch, I.R.; Leder, P.  
 Cell 29, 691-699, 1982  
 A: Title: Duplication and deletion in the human immunoglobulin epsilon genes.  
 A: Reference number: A90824; MUID: 83001945; PMID: 6288268  
 A: Accession: A90824  
 A: Molecule type: DNA  
 A: Residues: 1-358, 'L', 360-428 <MAX>  
 A: Cross-references: GB: J00222; NID: g184755  
 A: Note: this sequence difference may be due to polymorphism  
 R: Benich, H.H.; Johansson, S.G.O.; Von Bahr-Lindstrom, H.  
 in Immediate Hypersensitivity: Modern Concepts and Developments, Bach, M.K., ed., pp. 1-  
 A: Reference number: A94418  
 A: Accession: A94418  
 A: Molecule type: protein  
 A: Residues: 'GAWTL', 6, 'X', 8-16, 'B', 18-43, 'B', 45-52, 55-92, 95-97, 'B', 99-121, 'B', 123, 'L', 1  
 A: Experimental source: myeloma protein Nd  
 R: Kenen, J.H.; Molgaard, H.V.; Houghton, M.; Dethyshire, R.B.; Viney, J.; Bell, L.O.; C  
 Proc. Natl. Acad. Sci. U.S.A. 79, 6661-6665, 1982  
 A: Title: Cloning and sequence determination of the gene for the human immunoglobulin ep  
 A: Reference number: A93933; MUID: 83065234; PMID: 6815656  
 A: Accession: B93933  
 A: Molecule type: mRNA  
 A: Residues: 1-40, 68-114, 427-428 <KEN>  
 A: Cross-references: GB: L00022; NID: g185035  
 R: Ikegama, S.  
 FEBS Lett. 224, 306-310, 1987  
 A: Title: Purification and characterization of a recombinant human IGE Fc-epsilon fragmen  
 A: Reference number: S02438; MUID: 88083554; PMID: 3121387  
 A: Accession: S02438  
 A: Status: nucleic acid sequence not shown  
 A: Molecule type: mRNA  
 A: Residues: 98-352 <IKE>  
 R: Zhang, K.; Max, E.E.; Cheah, H.K.; Saxon, A.

U. Biol. Chem. 269, 456-462, 1994  
 A>Title: Complex alternative RNA splicing of epsilon-immunoglobulin transcripts produces  
 A:Reference number: A53116; MUID:94103254; PMID:8276835  
 A:Accession: A53116  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 320-428 <ZH2>  
 A:Experimental source: myeloma U266-derived cell line AF-10  
 A>Note: sequence extracted from NCBI backbone (NCBIN:141701, NCBIPI:141702)  
 R:Hellman, L.  
 Eur. J. Immunol. 23, 159-167, 1993  
 A>Title: Characterization of four novel epsilon chain mRNA and a comparative analysis of  
 A:Reference number: A46536; MUID:93122085; PMID:8419166  
 A:Accession: C46536  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 382-426 <HEL>  
 A:Cross-references: GB:S55273; NID:G263166; PIDN:AB24857.1; PID:G263167  
 A:Experimental source: B cell myeloma U-266  
 A>Note: sequence extracted from NCBI backbone (NCBIP:125297)  
 A:Accession: D46536  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 382-391 <HE2>  
 A:Cross-references: GB:S55276; NID:G263168; PIDN:AB24858.1; PID:G263169  
 A:Experimental source: B cell myeloma U-266  
 A>Note: sequence extracted from NCBI backbone (NCBIP:125299)  
 A:Accession: A46536  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 401-428 <HE3>  
 A:Cross-references: GB:S53497; NID:G263162; PIDN:AB24855.1; PID:G263163  
 A:Experimental source: B cell myeloma U-266  
 A>Note: sequence extracted from NCBI backbone (NCBIP:123483)  
 C:Genetics:  
 A:Gene: GDB:IGHE  
 A:Cross-references: GDB:119335; OMIM:147180  
 A:Map position: 14q32.33-14q32.33  
 A:Introns: 1/1; 104/1; 211/1; 319/1  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin  
 F:22-87/Domain: immunoglobulin homology <IM1>  
 F:128-195/Domain: immunoglobulin homology <IM2>  
 F:232-301/Domain: immunoglobulin homology <IM3>  
 F:338-407/Domain: immunoglobulin homology <IM4>  
 F:14/Disulfide bonds: interchain (to light chain) #status predicted  
 F:15-105,29-85,135-193,239-299,345-405/Disulfide bonds: #status predicted  
 F:21,49,99,146,252,275/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:121,209/Disulfide bonds: interchain (to heavy chain) #status predicted

Query Match 10.4%; Score 356; DB 1; Length 428;  
 Best Local Similarity 26.6%; Pred. No. 1e-13;  
 Matches 118; Conservative 69; Mismatches 154; Indels 102; Gaps 18;

QY 267 TQDRL-----QMGKLPRLHLT-----LFOALPOYAGSGL--TLALPAKTKGL 308  
 DB 3 TQSSVPEPLTCCCKNIPSNATSVTLGLATGYFPEPVVWTVDTGSLNGTITLTPATLTL 62  
 QY 309 ---HQEVNLVVMRATQLOKNTLCEV-WGPTSPKMLSLKLENKAKYSKRE---KPVWVL 361  
 DB 63 SGHATISLTLVSGAMAKOMFTCRVAHTPSTDM-----VNKKFPCSNDFTPTPKYL 117  
 QY 362 N-----DEAGMOCILS-----DSGOVLTESNIKVLPTWSTPVPCCAPE 400  
 DB 118 QSSCDGGHFPPTIOLCLVSGYTPGTINTWLTEDGQVMDVD---LSTASTQEGELAS 173  
 QY 401 PKS-----CDKHTC-----PELLGSGVFLPPPKKDTL 430  
 DB 174 TQSELTLGSKMLSDRYTCOVTVQGHTEFEDSTKCADSNR---GVSAYLSRPSPD-L 229  
 QY 431 MISRTPEVTGVVVDVSHEDPEVKFNWVYDVGEVNAKTKPREEQYNSYRVRVSLVTLVHQ 490

DB 230 FIRSPITTCIVDLAPSKGTVNLTWGRASGKPVNHSRREKQRNGTILVTSTLPVGR 289  
 QY 491 DMLNGKVKCKVSKALPAPEKTISSAKQKQRPQVYTL-----PSRDELTKNOVSLT 545  
 DB 290 DWIEGETYQCRVHPHPLRALMRSTTKTSGRAAPEVYATPPEWPSRDKRT-----LA 344  
 QY 546 CLVGFYPSDIAYEWESN--GQPENNYKTPPVLDSDGSPFLYSKLTVDKSRMOQGVNFS 603  
 DB 345 CLINFMPEDISVQMLNEVQLPDAKISTTQPRKTKSGFFVSRLEVTAEWQKDBFI 404  
 QY 604 CSVMEALHNHYT-QKSLSLSPG 625  
 DB 405 CRAVHEAASPSQTVQRAVSNVPG 427

RESULT 63  
 MIRM  
 Ig mu chain C region, membrane-bound form - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 04-Dec-1986 #sequence\_revision 30-Jun-1991 #text\_change 23-Aug-1997  
 C:Accession: A02165; A02164  
 R:Bernstein, R.E.; Alexander, C.B.; Reddy, E.P.; Mage, R.G.  
 U. Immunol. 132, 490-495, 1984  
 A>Title: Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain of V-Ha2  
 A:Reference number: A02164; MUID:84088930; PMID:6418803  
 A:Contents: a2 allotype  
 A:Accession: A02165  
 A:Molecule type: mRNA  
 A:Residues: 439-479 <BE2>  
 A:Accession: A02164  
 A:Molecule type: mRNA  
 A:Residues: 1-438, 'GKPTLYNSLIMSDFASTCY' <BE1>  
 A>Note: The sequence of residues 1-438 was assumed to be identical with the correspondi  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob  
 F:21-92/Domain: immunoglobulin homology <IM1>  
 F:130-202/Domain: immunoglobulin homology <IM2>  
 F:242-310/Domain: immunoglobulin homology <IM3>  
 F:349-420/Domain: immunoglobulin homology <IM4>  
 F:439-479/Domain: carboxyl-terminal <CTS>  
 F:14/Disulfide bonds: interchain (to light chain) #status predicted  
 F:28-90,137-200,249-308,356-418/Disulfide bonds: #status predicted  
 F:46,114,212,261,277,284/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:219/Disulfide bonds: interchain (to heavy chain) #status predicted  
 F:236/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.4%; Score 355.5; DB 1; Length 479;  
 Best Local Similarity 24.8%; Pred. No. 1.3e-13;  
 Matches 122; Conservative 81; Mismatches 176; Indels 113; Gaps 21;

QY 223 FPL-----AFTVEKLTGSGELMMQAEKSSKSMITFDLKNKEVSVKRTQDPLQMGKK 277  
 DB 9 YPLVSCGALTDGNLVAMGLADFLPSSVTFPSH-SRK-NNSEISSRTVTFPVVKGDK 66  
 QY 278 -----LPLHDTLPOAL-----POYA 292  
 DB 67 YMATSGVLVPSKDVLTQTEEYLVCKVQHSNSNDRLRFVDSDELPPNVSVFIPRDSFS 126  
 QY 293 GSGNLTALPAK-TGKLHQEVNLVVMRATQ-----LQKULTGEWQPTSPKMLSLKLE 345  
 DB 127 GSGTRSKRLICQATGSPKQISVSWLRDQGVESGVLTKEVBEATKAGPATFISIMLT 186  
 QY 346 NKEAKVSKREKPVVNLNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVPCCAPEPCSD 405  
 DB 187 ITESD-----WL-----SGLVTCRVDRHG-IFPKNVMSSECSL-----TPSP----- 225  
 QY 406 KHTCPPELLGSPVFLPPPKKDTLMIISRTPEVTGVVVDVSHEDPEVKFNWVYDVGEVNH 465  
 DB 226 -----GIQVFPIAPSFADT-FLSKSARLICLVTLDTTYG-SLNTISM-----ASHN 268

```

QY 466 AKT-----KPREQNSTYRVSVLTLYVHODMLNGEKYCKVSNKALPAPIETKISKAG 520
DB 269 GKRLDTHMTESHHPNATFSAEASVCAEDWESGQFTCTVHADLPPLKATIKSKRE 328
QY 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWEMESNGOP--ENNYKTPPV 576
DB 329 VAGHPRAVYVLPAPAREGLVRESATYTCVKGSPADVPVQMOGRQPLSSDKYVTSARA 388
QY 577 LD--SDGSFPLYSKLTVDKSRMOQGVFSCSVMEALHNHYTKSLSPGLDLETCAE 634
DB 389 PEPOAGLVFTYHSTLTVTEEDMNSGETFTCVGHEALPHWATERVYDKS-----TEGEV 443
QY 635 AODELDGLMTT 646
DB 444 AEEEGFENIMWT 455

RESULT 64
136948
Ig epsilon-chain - chimpanzee (fragment)
C/Species: Pan troglodytes (chimpanzee)
C/Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 21-Jan-2000
C/Accession: I36948
R/Sakoyama, Y.; Hong, K.
Proc. Natl. Acad. Sci. U.S.A. 84, 1080-1084, 1987
A/Title: Nucleotide sequences of immunoglobulin epsilon genes of chimpanzee and orangutan
A/Reference number: I36948; MUID:87147196; PMID:3103123
A/Accession: I36948
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-426 <RES>
A/Cross-references: GB:M15398; NID:g176797; PIND:AAA3416.1; PID:g176798
C/Genetic:
A/Intons: 103/1; 209/1; 317/1
C/Superfamily: immunoglobulin C region; immunoglobulin homology
F/336-405/Domain: immunoglobulin homology <IMM>

Query Match 10.4%; Score 353.5; DB 2; Length 426;
Best Local Similarity 26.4%; Pred. No. 1.5e-13;
Matches 110; Conservative 68; Mismatches 162; Indels 77; Gaps 15;

QY 228 TVEKLTGSGELMW-----QAERASSKSMI--TFDLKNKEVSRYKVTQDPLQWKK 277
DB 67 TISLITVSGA--NAKQMFTRVATPSTDMVNDKTPSVCSRDPFVYKVLQSSCDGGH 124
QY 278 LPLHLTLPOALPOYA--GSGNLTLALAKTKLHQEVNLVVMRAITQLOKXLTCEWGPSP 336
DB 125 FPPITQLCLVSGYTPGTINITWLEDOV---MDVDLSTASATQEGE-----LASTGS 174
QY 337 KLMISLLENKEAKVSKREKRPVWVNLPEAGMOCILSDSQVLLSEINIKVLPWSTPVP 396
DB 175 ELTLISQK-----HMLSD--RTYTCQVTVYQGTREDSTK-----C 207
QY 397 PAPEPKSCDKHTHPCPELLGSPVFLPPKPRDTLMISRTPEVTCVVVDVSHEDPEVKFN 456
DB 208 ADSNPR-----GVSAFLSRPSFPD-LFIKSPITICLVVDLAPSKGTALVT 253
QY 457 YVDGVEVNAKTKRREQVNSTYRVSVLTLYVHODMLNGEKYCKVSNKALPAPIETKIS 516
DB 254 SRASGKRVNSTRKQERQKRGITLVSTLPVGRDWMIEGEVYQCRVTHPHLPALVSTT 313
QY 517 KAKQPREPOVYTL-----PPSRDELTKNOVSLTCLVKGFPSPDIWEMESN--GQPENN 569
DB 314 KTSPPRAAPAYATPAREGSPSRKRT-----LACLQNTMPEDISVQMLHNEVOLP 368
QY 570 YKTPPVLDSGSPFLYSKLTVDKSRMOQGVFSCSVMEALHNHYT--OKSLSPG 625
DB 369 HSTTQPKTKSGGFVSRLEVTAEWQKDEFCRAVNHAPSQVTVQTVSVNPG 425

RESULT 65
MHMUBT
Ig mu heavy chain disease protein (Boc) - human

```

```

C/Species: Homo sapiens (man)
C/Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jul-1999
C/Accession: A02163
R/Barnikol-Watanabe, S.; Mihaesco, E.; Mihaesco, C.; Barnikol, H.U.; Hilschmann, N.
Hope-Seyler's Z. Physiol. Chem. 365, 105-118, 1984
A/Title: The primary structure of mu-chain-disease protein BOT. Peculiar amino-acid seq
A/Reference number: A02163; MUID:84184186; PMID:6425189
A/Accession: A02163
A/Molecule type: protein
A/Residues: 1-391 <BAR>
A/Comment: This protein has no V region homology or CH1 region.
C/Genetic:
A/Genes: GDB:IGHM
A/Cross-references: GDB:120086; OMIM:147020
A/Map position: 14q32.33-14q32.33
A/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin; transmembrane pr
F/1-41/Domain: pre-C <VAR>
F/43-391/Domain: Ig mu chain C region, secreted form <IGH>
F/65-137/Domain: immunoglobulin homology <IMM1>
F/175-243/Domain: immunoglobulin homology <IMM2>
F/282-353/Domain: immunoglobulin homology <IMM3>
F/147,210,217,378/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.3%; Score 352.5; DB 1; Length 391;
Best Local Similarity 27.8%; Pred. No. 1.5e-13;
Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;

QY 332 GPTSPRLMLSLLENKEAKVSKREKRPVWVNLPEAGMOCILSDSQVLLSEINIKVL--PT 389
DB 110 GPTTYKVTSTLTIKSD-----ML---GQSMFTCRVDHRG-LTRQVNASMGCPD 155
QY 390 WSTPVPAPAPKSCDKHTHPCPELLGSPVFLPPKPRDTLMISRTPEVTCVVVDVSHD 449
DB 156 QDTAI-----RVFAIPPS--FASIFLYSTKTLCLVTDLTVD 191
QY 450 PEYKFMVYDGVENNAKTKPREQVNSTYRVSVLTLYVHODMLNGEKYCKVSNALPA 509
DB 192 -SVTISWTRQDEAVAKTHTNISEHPNATFSAEASVCAEDWESGQFTCTVHADLP 250
QY 510 PIEKTSKAKGP-REPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWEMESNGOP- 566
DB 251 PLKQTSRPRKGVALHRPDVYLLPPAREQLNLRESATITCLVGFSPADVFMQMOGRQPL 310
QY 567 -ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGVFSCSVMEALHNHYTKSLSPG 623
DB 311 SPEKYTSAPMPPEPQAPGRYFAHSILTVSEEMWGTGETYTCVVAHEALPNRYVTRVDS 370
QY 624 PG 625
DB 371 TG 372

RESULT 66
MHMUM
Ig mu chain C region, membrane-bound splice form - human
C/Species: Homo sapiens (man)
C/Date: 29-Jan-1993 #sequence_revision 23-Aug-1997 #text_change 22-Jun-1999
C/Accession: S16510; S09357; S16556; B26243; A02167
R/Dorai, H.; Gillies, S.D.
Nucleic Acids Res. 17, 6412, 1989
A/Title: The complete nucleotide sequence of a human immunoglobulin genomic C-mu gene.
A/Reference number: S09357; MUID:89366690; PMID:2505237
A/Accession: S16510
A/Molecule type: DNA
A/Residues: 433-473 <DOR1>
A/Cross-references: EMBL:X14939
A/Note: the authors translated the codon AAC for residue 445 as Met
A/Note: the sequence of residues 1-432 was assumed to be identical with the correspondi
A/Molecule type: DNA
A/Residues: 1-432; GKPLVNVSLVMSDPTAGTCY' <DOR2>
A/Cross-references: EMBL:X14940

```





```

Oy      309 HOEVLV-----WKATQJOKNICEWGFSPKYMILKLEKNEKAVSREKPPVVLN 362
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      111 NPNVNVFVPPRDGFSGPAPRKSKLICEATNFPKPIYMSLKDGKLYESGTTDPVITEN 170

Oy      363 ---PEA-----GMOCLSDSGQVLLEBSNIKVLPTWSTPVPBPAPBP 401
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      171 KQSTPQTYKVISITLISEIDMLNMYTCRVDBHGLTFLK----- 210

Oy      402 KSCDKHTHCBELIGPS---VFLPPPKPKDTLMTSRTPEVTCVVDVSHEDPEVKENWY 457
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      211 ---NVSSTC---AASBSTDILFTTIPPSFAD-IFPSKANLTCVSNLATIVE-TLNTSMA 262

Oy      458 VDGVEVHNNAKTKREREQNYSYTRVSVLTVLHODVNLKEKCKVSKNALPAPLEKITSK 517
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      263 SSGSEPLETKIKIMESHNGTFSANGVASVCEJDMNNKKEPCVTHRDDLPSPCKFISK 322

Oy      518 AKQBPB-----POVYTLPRSPDELT-KNQVSLTCLVGFPSDIAVEMESNGO--PENN 569
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      323 ---PNEVNHGHPAVYLLPPAREQNLNRASAVTCLVGFSPADISVOMLORGGLPEOK 378

Oy      570 YKTTTPVL D--SDGSFFLYSKLTVPKSRKQOQNVGSCVMEHALNHNHTQKSLSPG 625
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      379 YVTSAPMEPEGAGCFYTHSILTVLEENNSSETYTCVGHGALPHLVTEKTVKSTG 436

```

```

RESULT 68
A24976
Ig mu chain C region (allele b) - mouse
C/Species: Mus musculus (house mouse)
C/Date: 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 21-Jan-2000
C/Accession: A24976
R/Schreier, P.H.; Quester, S.; Botwell, A.
Nucleic Acids Res. 14, 2381-2389, 1986
A/Title: Alleotypic differences in murine mu-genes.
A/Reference number: A24976; MUID:86176735; PMID:3083402
A/Accession: A24976
A/Molecule type: mRNA
A/Residues: 1-455 <SCH>
A/Cross-references: GB:X03690; NID:952381; PIDN:CAA27326.1; PID:952382
A/Experimental source: strain C57BL/6
A/Note: the authors translated the codon AAG for residue 65 as Leu
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/346-417/Domain: immunoglobulin homology <IMM>

Query Match      10.2%; Score 348; DB 2; Length 455;
Best Local Similarity 22.7%; Pred. No. 3.3e-13;
Matches 122; Conservative 77; Mismatches 179; Indels 160; Gaps 18;

QY      144 ESPRESSSVOCRRP-RKNIQGGKTLVSQLELDGSGTWCTVLQNKQKVEFKIDIVL 202
      :::::|:::|:::|
DB       3 QSFVVFPLVACESLSQKNI.VAMGCLARDFLPSTISFTWN----- 43

QY      203 AFQKASIVYKKEGEQVEFPLAFTVEKLTGSGELMWQABRASSKSMITFDLKNKEVS 262
      :::|:::|:::|
DB       44 -----YQNTNEVIQ-----GIRFPLRTQGGKYLA 68

QY      263 VKRVTDPRKLN-----GKKLPIHLTLPOLPOYAGSGNLTALAEKTKL 308
      :::|:::|:::|
DB       69 TSOVLISFKSLIEGDEVLVCKIHGGKNDLHPIP-----AVAEW 110

QY      309 HOEVNLVV-----WRATQLOKNLICEWGPSPRYLMLSLKLENKEAVSRKRPVWVLN 362
      :::|:::|:::|
DB      111 NPNNVVFPPRDGSGPAPRSKLLCEATNFTPKRTIYSLWKDGKLVESGFTTDPVTIEN 170

QY      363 ---PEA-----GMQCLLSDSQVLLESNIKVLPTWSPTPVPCAPAP 401
      :::|:::|:::|
DB      171 KGSFQTVKVIISTLTISEIDMLNLVNVCYRVDRGLTEFK----- 210

QY      402 KSCDKTHICPELGGPS---VELPRPKQKLTMLSRPEVTCVAVVDVSHEDPEVKRWY 457
      :::|:::|:::|
DB      211 ---NVSSYTC---AASPSTDLITFTLPSPAD-IFLSKSNANLTCLVSNLATYE-TLNISWA 262

```

```

Oy      458 VDGEVHNHAKKPREEGNSTYRVRVSVLTVTLODMJNGEYKCYKSNKLPPIETKIS 517
Db      263 SÖSGEPLTKIKIMESHNGTFSAKGAVSVCYEDNNRKEPCTYTHRLPSPÖKFISK 322

Oy      518 AKGPRE-----POVYTLPSRDELT-KNOVSLTCLVGFPSDIAVEMESNGO--PENN 569
Db      323 -----PNEVHKHPRAVYTLPRAREQUNLRESATVTCLVNGFSPADISVQMLQSGQLLPÖEK 378

Oy      570 YKTPRPVLD--SDGSFLLYSLKLTVDKSHMOQGNVESCSVMHEALHNHYTÖKSLSLSPG 625
Db      379 YVTGAPMEPRPARGCYFTHSILTYTEEMNSSETTYTCVGHGHALPHLYTERVYDSTG 436

```

RESULT 69

MMHU

Ig mu chain C region, secreted splice form - human

C:Species: Homo sapiens (man)

C:Date: 29-Jul-1991 #sequence, revision 23-Aug-1997 #text, change 22-Jun-1999

C:Accession: S09357; S16556; S37767; A26243; A26244; I37749; I37750; A02162; B02162

R:Donat: H.; Gillies, S.D.

Nucleic Acids Res. 17, 6412, 1989

A:Title: The complete nucleotide sequence of a human immunoglobulin genomic C-mu gene.

A:Reference number: S09357; MUID:89366690; PMID:2505237

A:Accession: S09357

A:Molecule type: DNA

A:Residues: 1-452 <DOR1>

A:Cross-references: EMBL:X14940

A:Note: the authors translated the codon AAT for residue 16 as Met

R:Donat: H.

submitted to the EMBL Data Library, April 1989

A:Reference number: S16556

A:Accession: S16556

A:Molecule type: DNA

A:Residues: 1-39, 'L', 41-452 <DOR2>

A:Cross-references: EMBL:X14940

R:Harindranath, N.; Donadel, G.; Sigounas, G.; Nockins, A.L.

Mol. Immunol. 30, 111-112, 1993

A:Title: Comparison of complete nucleotide sequence of the human IgM heavy chain constant

A:Reference number: S37767; MUID:93109369; PMID:8417370

A:Accession: S37767

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-190, 'S', 192-414, 'V', 415-452 <HAR>

A:Cross-references: EMBL:X67292; NID:G38405; PIDN:CAA47708.1; PID:G38406

R:Rabbits, T.H.; Forester, A.; Milestein, C.P.

Nucleic Acids Res. 9, 4509-4524, 1981

A:Title: Human immunoglobulin heavy chain genes: evolutionary comparisons of C-mu, C-delta

A:Reference number: A26243; MUID:82059479; PMID:6795553

A:Accession: A26243

A:Molecule type: DNA

A:Residues: 1-17, 105-186; 200-259; 296-322; 339-416, 'D', 418-452 <RAB>

A:Cross-references: GB:K01310; NID:G184715

R:Dooley, T.W.; Devono, J.; Croce, C.M.

Proc. Natl. Acad. Sci. U.S.A. 77, 6027-6031, 1980

A:Title: Cloning and partial nucleotide sequence of human immunoglobulin mu chain cDNA

A:Reference number: A26244; MUID:81077306; PMID:6777778

A:Accession: A26244

A:Molecule type: mRNA

A:Residues: 298-386; 436-452 <DOL>

A:Cross-references: GB:J00257; NID:g185053; PIDN:AAA53508.1; PID:g185056; GB:J00258; NID:J00259

R:Takahashi, N.; Nakai, S.; Honjo, T.

Nucleic Acids Res. 8, 5983-5991, 1980

A:Title: Cloning of human immunoglobulin mu gene and comparison with mouse mu gene.

A:Reference number: I37748; MUID:81124312; PMID:6450943

A:Accession: I37749

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 347-370 <TAK1>

A:Cross-references: EMBL:V00562; NID:G33448; PIDN:CAA23825.1; PID:G929649

A:Accession: I37750

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 433-452 <TAK2>



A:Cross-references: EMBL:V009563; NID:933454; PIDD:CAA2826.1; PID:G9825664  
R:Mañesco, E.; Barnikol-McNane, S.; Barnikol, H.U.; Mihaesco, C.; Hilschmann, N.  
Eur. J. Biochem. 111, 275-286, 1980  
A:Title: The primary structure of the constant part of mu-chain-disease protein BOT.  
A:Reference number: A02162; MUID:81066716; PMID:6777162  
A:Contents: Mu-chain-disease protein Bot and revisions to sequence of Gal  
A:Accession: A02162  
A:Molecule type: protein  
A:Residues: 1-17, 'BPS', '22', 'T', '24-82', 'N', '84-90', 'Z', '92-93', 'B', '95', 'B', '97-144', 'E', '146-162', 'B', '163-164', 'E', '165-166', 'B', '167-168', 'E', '169-170', 'B', '171-172', 'E', '173-174', 'B', '175-176', 'E', '177-178', 'B', '179-180', 'E', '181-182', 'B', '183-184', 'E', '185-186', 'B', '187-188', 'E', '189-190', 'B', '191-192', 'E', '193-194', 'B', '195-196', 'E', '197-198', 'B', '199-200', 'E', '201-202', 'B', '203-204', 'E', '205-206', 'B', '207-208', 'E', '209-210', 'B', '211-212', 'E', '213-214', 'B', '215-216', 'E', '217-218', 'B', '219-220', 'E', '221-222', 'B', '223-224', 'E', '225-226', 'B', '227-228', 'E', '229-230', 'B', '231-232', 'E', '233-234', 'B', '235-236', 'E', '237-238', 'B', '239-240', 'E', '241-242', 'B', '243-244', 'E', '245-246', 'B', '247-248', 'E', '249-250', 'B', '251-252', 'E', '253-254', 'B', '255-256', 'E', '257-258', 'B', '259-260', 'E', '261-262', 'B', '263-264', 'E', '265-266', 'B', '267-268', 'E', '269-270', 'B', '271-272', 'E', '273-274', 'B', '275-276', 'E', '277-278', 'B', '279-280', 'E', '281-282', 'B', '283-284', 'E', '285-286', 'B', '287-288', 'E', '289-290', 'B', '291-292', 'E', '293-294', 'B', '295-296', 'E', '297-298', 'B', '299-300', 'E', '301-302', 'B', '303-304', 'E', '305-306', 'B', '307-308', 'E', '309-310', 'B', '311-312', 'E', '313-314', 'B', '315-316', 'E', '317-318', 'B', '319-320', 'E', '321-322', 'B', '323-324', 'E', '325-326', 'B', '327-328', 'E', '329-330', 'B', '331-332', 'E', '333-334', 'B', '335-336', 'E', '337-338', 'B', '339-340', 'E', '341-342', 'B', '343-344', 'E', '345-346', 'B', '347-348', 'E', '349-350', 'B', '351-352', 'E', '353-354', 'B', '355-356', 'E', '357-358', 'B', '359-360', 'E', '361-362', 'B', '363-364', 'E', '365-366', 'B', '367-368', 'E', '369-370', 'B', '371-372', 'E', '373-374', 'B', '375-376', 'E', '377-378', 'B', '379-380', 'E', '381-382', 'B', '383-384', 'E', '385-386', 'B', '387-388', 'E', '389-390', 'B', '391-392', 'E', '393-394', 'B', '395-396', 'E', '397-398', 'B', '399-400', 'E', '401-402', 'B', '403-404', 'E', '405-406', 'B', '407-408', 'E', '409-410', 'B', '411-412', 'E', '413-414', 'B', '415-416', 'E', '417-418', 'B', '419-420', 'E', '421-422', 'B', '423-424', 'E', '425-426', 'B', '427-428', 'E', '429-430', 'B', '431-432', 'E', '433-434', 'B', '435-436', 'E', '437-438', 'B', '439-440', 'E', '441-442', 'B', '443-444', 'E', '445-446', 'B', '447-448', 'E', '449-450', 'B', '451-452', 'E', '453-454', 'B', '455-456', 'E', '457-458', 'B', '459-460', 'E', '461-462', 'B', '463-464', 'E', '465-466', 'B', '467-468', 'E', '469-470', 'B', '471-472', 'E', '473-474', 'B', '475-476', 'E', '477-478', 'B', '479-480', 'E', '481-482', 'B', '483-484', 'E', '485-486', 'B', '487-488', 'E', '489-490', 'B', '491-492', 'E', '493-494', 'B', '495-496', 'E', '497-498', 'B', '499-500', 'E', '501-502', 'B', '503-504', 'E', '505-506', 'B', '507-508', 'E', '509-510', 'B', '511-512', 'E', '513-514', 'B', '515-516', 'E', '517-518', 'B', '519-520', 'E', '521-522', 'B', '523-524', 'E', '525-526', 'B', '527-528', 'E', '529-530', 'B', '531-532', 'E', '533-534', 'B', '535-536', 'E', '537-538', 'B', '539-540', 'E', '541-542', 'B', '543-544', 'E', '545-546', 'B', '547-548', 'E', '549-550', 'B', '551-552', 'E', '553-554', 'B', '555-556', 'E', '557-558', 'B', '559-560', 'E', '561-562', 'B', '563-564', 'E', '565-566', 'B', '567-568', 'E', '569-570', 'B', '571-572', 'E', '573-574', 'B', '575-576', 'E', '577-578', 'B', '579-580', 'E', '581-582', 'B', '583-584', 'E', '585-586', 'B', '587-588', 'E', '589-590', 'B', '591-592', 'E', '593-594', 'B', '595-596', 'E', '597-598', 'B', '599-600', 'E', '601-602', 'B', '603-604', 'E', '605-606', 'B', '607-608', 'E', '609-610', 'B', '611-612', 'E', '613-614', 'B', '615-616', 'E', '617-618', 'B', '619-620', 'E', '621-622', 'B', '623-624', 'E', '625-626', 'B', '627-628', 'E', '629-630', 'B', '631-632', 'E', '633-634', 'B', '635-636', 'E', '637-638', 'B', '639-640', 'E', '641-642', 'B', '643-644', 'E', '645-646', 'B', '647-648', 'E', '649-650', 'B', '651-652', 'E', '653-654', 'B', '655-656', 'E', '657-658', 'B', '659-660', 'E', '661-662', 'B', '663-664', 'E', '665-666', 'B', '667-668', 'E', '669-670', 'B', '671-672', 'E', '673-674', 'B', '675-676', 'E', '677-678', 'B', '679-680', 'E', '681-682', 'B', '683-684', 'E', '685-686', 'B', '687-688', 'E', '689-690', 'B', '691-692', 'E', '693-694', 'B', '695-696', 'E', '697-698', 'B', '699-700', 'E', '701-702', 'B', '703-704', 'E', '705-706', 'B', '707-708', 'E', '709-710', 'B', '711-712', 'E', '713-714', 'B', '715-716', 'E', '717-718', 'B', '719-720', 'E', '721-722', 'B', '723-724', 'E', '725-726', 'B', '727-728', 'E', '729-730', 'B', '731-732', 'E', '733-734', 'B', '735-736', 'E', '737-738', 'B', '739-740', 'E', '741-742', 'B', '743-744', 'E', '745-746', 'B', '747-748', 'E', '749-750', 'B', '751-752', 'E', '753-754', 'B', '755-756', 'E', '757-758', 'B', '759-760', 'E', '761-762', 'B', '763-764', 'E', '765-766', 'B', '767-768', 'E', '769-770', 'B', '771-772', 'E', '773-774', 'B', '775-776', 'E', '777-778', 'B', '779-780', 'E', '781-782', 'B', '783-784', 'E', '785-786', 'B', '787-788', 'E', '789-790', 'B', '791-792', 'E', '793-794', 'B', '795-

```

Db      433 G 433

RESULT 70
EMMS
Ig epsilon chain C region (version 2) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Aug-1996
C:Accession: A02145
R:Shida, N.; Ueda, S.; Hayashida, H.; Miyata, T.; Honjo, T.
EMBO J. 1, 1117-1123, 1982
A>Title: The nucleotide sequence of the mouse immunoglobulin epsilon gene: comparison w
A:Reference number: A50966; MUID:84236092; PMID:6329728
A:Accession: A02145
A:Molecule type: DNA
A:Residues: 1-423 <ISH>
A>Note: the sequence was determined from the germ-line gene
C:Genetics:
A:Introns: 91/1; 199/1; 307/1
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:15-77/Domain: immunoglobulin homology <IM1>
F:115-183/Domain: immunoglobulin homology <IM2>
F:120-288/Domain: immunoglobulin homology <IM3>
F:325-396/Domain: immunoglobulin homology <IM4>
F:23-75, 122-181, 227-286, 332-394/Disulfide bonds: #status predicted
F:43, 84, 167, 239, 265, 417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      10.1%; Score 346.5; DB 1; Length 423;
Best Local Similarity 26.5%; Pred. No. 3,76-13;
Matches 118; Conservative 74; Mismatches 154; Indels 99; Gaps 21;

Qy      223 FPLAFYEKLTGSELMWQAERASSSKSWITFDLKNK--VSVRKYVDPRQLQNGKULPL 280
      |||||
Db      29 FPNPVTYV-----WYSD--SLNMTVPFALGSELKVTISQVTSWGK--SAKNFTC 75

Qy      281 HLTEFOALPOYAGSGNLTLEAKTGKHOEVLVNR---ATOLQNLTCFVWGPSTP 336
      |||||
Db      76 HVTHPSPFNE-----SRITLVRPVTHSLSPWMSIHRCPNAPHSTIQLYCFYGHILN 130

Qy      337 KMLSLKLENK-----AKVSRKRPVAVLNPAGMWQ-----CLLSDSQVL 379
      |||||
Db      131 DVSVSMIMDDREIDTLAQTVLKEEGKLASTGSKLNTIEQOMSESTFTCRVTSQGVY 190

Qy      380 LESNINKVLPTSTVPCAPAPKSCDKTHCPPELLGGPSVFLFPPPKKDTLMISRTVEY 439
      |||||
Db      191 LAHTRR-----CPDHEPR-----CAITVLLPPSPLD--LYNGAPLTL 226

Qy      440 CVVADV--SHDEPEYKFN-----WYVDGVEVNAKATKPREEQNSTYRVVSLTV 487
      |||||
Db      227 CLVVDLESEKKNVNTWNOEKTTSVASQWY--TKHN-----NATITSTSLIPV 273

Qy      488 LHODMLNGEKKYCKVSKKALPAPIEKTSISKAK--GQPREPOVYTLPPSRDELTKNQVSLTC 546
      |||||
Db      274 VAKOMIEGYGVQCVVDRPDPFKPIVRSITLIPQVQSRGAPDEVYVPPPEEE--SEDKRTLTC 332

Qy      547 LVKGFYSDDIAVNESNGQPENN--YKTTTPVLDSDS---FLYSKLTYDKSKWQGNV 601
      |||||
Db      333 LIQGFPEPDISVQWLGQGLKLSNQSHSTTP--LKSNGSNOGFFLFSRLLEVAKTLMTRQK 391

Qy      602 FSCVMGHEALHN--HYTOKSLSLSPG 625
      |||||
Db      392 FTCOVIEHALQPKRKLEKTIISTSLG 416

RESULT 71
EMHY
Ig mu chain C region - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Aug-1996

```



Best Local Similarity 24.8%; Pred. No. 4.4e-13;  
Matches 117; Conservative 78; Mismatches 168; Indels 108; Gaps 20;

```

Qy 223 FPL-----AFTVEKLTGSGELMWOERASSKSWITPDLKNEKVSVKRVTDPKLOQK 277
    |||
Db 9 YPLSCGALTDGNLVAMGCLARDPLPSSVTFSSW-SFK-NNSEISSRTVTRFPVKKGD 66
    |||
Qy 278 -----LPLHLPLPOL-----POYA 292
    |||
Db 67 YMATSOVLVPSKDVLOQTEERYLVCKVQHSNNDLRVSFPVDSLEPPVNSVFIIPRDSFS 126
    |||
Qy 293 GSGNLTLLLEAK-TGKLHOEVNLVVMRATQ-----LOKNLTCEVGPSTPKMLSLKLE 345
    |||
Db 127 GSGRKSLICQAGTFSPKQISVSWLRDQGVESGVLTKEPAEATKAGAPRTFISSLT 186
    |||
Qy 346 NKEAKVSKREKRPVWVNLPEAGMOCCLSDSQVLLSNIKVLPTWSTPVPCPAPEKSCD 405
    |||
Db 187 ITESD-----WL-----SQSLYTCRVDRHG-IFEDKNVSSSECSST-----TPSP----- 225
    |||
Qy 406 KTHCPPELLGSPVFLPPPKKDTLMTSRTEVTCVVVDVSHEDPEVKFNKYVDGVEVHN 465
    |||
Db 226 -----GIQFPPIAPSPADT-FLSKSARLCLVTDLTYYG-SLNIISW-----ASHN 268
    |||
Qy 466 AKT-----KPREEQNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEXTISKAKG 520
    |||
Db 269 GKALDTHNITESHFNATFSAMGEMSVCAEDMESGEPTCTVTYHADLPFLKHTISKSRE 328
    |||
Qy 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIWEMESNGQP--ENNYKTPPV 576
    |||
Db 329 VAKHPAYVLPAREQLVRESATVTCVKGFSPADVFVQMOQRGLSDKVTYSAPA 388
    |||
Qy 577 LD--SDGSFLLYSLKTLVDKSRMOQGVNFGSVNHEALHNHTOKSLSLSPG 625
    |||
Db 389 PEQAPGLYFTHSTLTYTEEDMNSGETFTCVAGHEALPHMWTERTVDKSTG 439
    |||

```

## RESULT 74

IS0731  
Ig heavy chain - nurse shark  
C/Species: Glinglyostoma citratum (nurse shark)  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C/Accession: IS0731  
R/Vazquez, M.; Mituki, N.; Flajnik, M.F.; McKinney, E.C.; Kaashara, M.  
Mol. Immunol. 29, 1157-1158, 1992  
A/Title: Nucleotide sequence of a nurse shark immunoglobulin heavy chain cDNA clone.  
A/Reference number: IS0731; MUID:92357056; PMID:1495502  
A/Accession: IS0731  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-577 <VAZ>  
A/Cross-References: GB:M92851; NID:G213264; PIDN:AAA50817.1; PID:G213265  
C/Genetic8:  
A/Genes: IGH  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
F:469-539/Domain: immunoglobulin homology <IMM>

Query Match 10.1%; Score 345.5; DB 2; Length 577;  
Best Local Similarity 23.9%; Pred. No. 6.3e-13;  
Matches 155; Conservative 110; Mismatches 256; Indels 127; Gaps 29;

```

Qy 10 LLLVLAQLALPAAQGNKV-----LGKKDTEVELTCAKSKSIQPMKSNQ--IKIL 62
    |||
Db 6 IFLSLALLLPVCGSEITLQPEARTGPGSLSTC-----KTRGFVLGSSSMYWRQV 60
    |||
Qy 63 GNCG-----SFLTKGPKLNDRADSRSLWDQGN--PPLIKMLKIEDSDTYICEV 111
    |||
Db 61 PGQGLEWIVYVYSSSMNNYPAIDRFPAK--DTSNIFALMBRSKIDDTAIYIC-- 115
    |||
Qy 112 EDQKEVQLLVGLTANSDTHLLOGQSITLLESPPGSSPSVQCRSPKGNIOGKTLV 171
    |||
Db 116 -----TRMSGYEYLGHSGYWGQGTWVTTATP--SSPLY-----GLVSSC 157
    |||
Qy 172 SQLELDQSGTWTCTVNLQNKVVEPKIDIVLAFQKASSIVYKKGGEQVPSFLAFYVK 231
    |||

```

Db 158 QQGNIDSSVYIGCLAM-----DYSPIV-----ASVTMKKGGLITGVQTPYSEVRN 203

```

Qy 232 LTGSGELMWOERASSKSWITPDLKNEKVS--VKRVTDPKLOMKKPLHLTPALP 289
    |||
Db 204 KKGTYTL-----SSQALIESPACDCQISCEVRHSGSKSTGM-----PCPDGFPALL 252
    |||
Qy 290 QVAGSGNLT-----LALFAKTKHGEVNLVVMRATQLOKNLTCEVGPSTPKMLSLKLE 345
    |||
Db 253 TVSSSEIEERKFAITVCSISDFRSK--SISVTMLKNGRSVDSGIFTSP-----VCEA 303
    |||
Qy 346 NKEAKVSKREKRPVWVNLPEAGMOCCLSDSQVLLSNIKVLPTWSTPVPCPAPEKSCD 405
    |||
Db 304 NGNSVTSRLRVPAEMFDAAVTC-----QVKK--EVIGSN-----ITGQVSECH 350
    |||
Qy 406 KTHCPPELLGSPVFLPPPKKDTLMTSRTEVTCVVVDVSHEDPEVKFNKYVDGVEVHN 465
    |||
Db 351 -----GYPAKILPP-PVEQVLLAEATVTLTCV--VSNLHSGVNFMTLOD----- 390
    |||
Qy 466 AKTKPREQYNS--TYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEXTISKAK-GQ 521
    |||
Db 391 EKLKSIADSGHSDGALSKLDISTEAWLSEVEECVNHQYLLPPLDSIHKEHEN 450
    |||
Qy 522 PREPOVYTLPPSRDELTKNO-VSLTCLVKGFPYPSDIWEMESNGQPEN--NYKTPPVLD 578
    |||
Db 451 PLERSVAVLPTTEHLSAQRFSLTCLVGRFRREIIVKKTINDKPYNPSNYKNTETVTA 510
    |||
Qy 579 SDG-SFFLYSLKTLVDKSRMOQGVNFGSVNHEALHNHTOKSLSLSPG 625
    |||
Db 511 SDNTSFYLSLLSIAEEMASGASVSCVGHGHALPLKINRTVDKSSG 558
    |||

```

## RESULT 75

IS2644  
Ig mu chain C region - rat (firearm)  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 20-Feb-1995 #sequence\_revision 30-Jan-1998 #text\_change 21-Jan-2000  
C/Accession: IS2644  
R/Parker, K.; Bugeon, L.; Soulliou, J.P.  
submitted to the EMBL Data Library, September 1992  
A/Reference number: IS2644  
A/Accession: IS2644  
A/Molecule type: mRNA  
A/Residues: 1-343 <PAR>  
A/Cross-References: EMBL:X68312; NID:956461; PIDN:CAA48392.1; PID:9818025  
A/Experimental source: spleen  
C/Genetic8:  
A/Map position: 6  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: immunoglobulin  
F:234-305/Domain: immunoglobulin homology <IMM>

Query Match 10.0%; Score 341.5; DB 2; Length 343;  
Best Local Similarity 29.5%; Pred. No. 5.6e-13;  
Matches 96; Conservative 59; Mismatches 129; Indels 41; Gaps 12;

```

Qy 323 QKNLTCEVWGPSTPKMLSLKENKAKVSKREKRPVWVNLPEAGMOCCLSDSQVLLS 382
    |||
Db 19 KSRLICATNFSPKQITVSWLODKPVKSFTTEPVTV--BAKGSRPQTYKIVISTLTITES 77
    |||
Qy 383 ---NIKV-----LPTWSTPVPCPAPEKSCDKTHCPPELLGSPVFLPPPKKDTL 430
    |||
Db 78 DWLNLNFTCRVDHRTGTLTFKKNVSTCAASPT-----DILAF-----IPPSFAD-I 124
    |||
Qy 431 MISTPEVTCVVVDVSHEDPEVKFNKYVDGVEVHNATKPREQYNSTYRVSVLTVLHQ 490
    |||
Db 125 FLTKSATLSGLVNTLAIYD-TLNISMSKSGEFLNTTKIMESHNPOTFPAVCAVASCME 183
    |||
Qy 491 DWLNGEKYCKVSNKALPAPIEXTISKAKQPRE-----POVYTLPPSRDEL-TKNQVSL 544
    |||
Db 184 DMDNRKSFVCTVTHRDLPSPQKFKISK--PNEVAKHPAVVYLLPPARQQLIRBSATV 239
    |||
Qy 545 TCLVKGFPYPSDIWEMESNGQP--ENNYKTPPVLD--SDGSFLLYSLKTLVDKSRMOQGN 600
    |||

```

Db 240 TCIVKFSFADIVQWLGKQPLSSDKYVTSAMPBPGAGLYFTHSILTVIEEANSSE 299  
 QY 601 VFSGSVMEALAHNYTKSLSPG 625  
 Db 300 TYTCVGVGHEALPHMVTERTVDKSTG 324

## RESULT 76

MHDG  
 Ig mu chain C region - dog (tentative sequence)  
 C/Species: Canis lupus familiaris (dog)  
 C/Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 31-Mar-2000  
 C/Accession: A93131; A94246; A02169  
 R/McCumber, L.J.; Capra, J.D.  
 Mol. Immunol. 16, 565-570, 1979  
 A/Title: The complete amino-acid sequence of a canine mu chain.  
 A/Reference number: A93131; MUID:80077682; PMID:117299  
 A/Contents: myeloma protein Moo  
 A/Accession: A93131  
 A/Molecule type: protein  
 A/Residues: 1-177 <MCC>  
 R/Wasserman, R.L.; Capra, J.D.  
 Science 200, 1159-1161, 1978  
 A/Title: Amino acid sequence of the Fc region of a canine immunoglobulin M: interspecies  
 A/Reference number: A94246; MUID:78180587; PMID:653360  
 A/Contents: Moo  
 A/Accession: A94246  
 A/Molecule type: protein  
 A/Residues: 178-450 <WAS>  
 C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin  
 F/20-89/Domain: immunoglobulin homology <IMM1>  
 F/126-196/Domain: immunoglobulin homology <IMM2>  
 F/234-302/Domain: immunoglobulin homology <IMM3>  
 F/341-412/Domain: immunoglobulin homology <IMM4>  
 F/430-450/Domain: carboxyl-terminal <CTS>  
 F/14/Disulfide bonds: Interchain (to light chain) #status predicted  
 F/27-87,133-194,348-410/Disulfide bonds: #status predicted  
 F/50,206,269,276,437/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F/211,449/Disulfide bonds: Interchain (to heavy chain) #status predicted  
 F/288/Disulfide bonds: Interchain (to mu chain in another subunit) #status predicted

Query Match 10.0%; Score 340; DB 1; Length 450;

Best Local Similarity 25.9%; Pred. No. 9,7e-13;

Matches 100; Conservative 76; Mismatches 142; Indels 68; Gaps 15;

QY 278 LPLHLTLPLQAL-----ROYAGSGN---LTLALEKTKGKLMHOENVLVMRATOLQKNLT-- 327  
 Db 102 LPWMLTLPPVSGFTIPRDAFPGSPKSQLICQSGSPQVWSL--RDGKQIESGVTIN 159  
 QY 328 -----CEWGPSTSEKMLSLLENKEAKVSKREKPVVNLPEAGMOCCLSDSGOVLLES 382  
 Db 160 EVZAZAKZSGPTTYKYVSMULTI-----QEDAL---SQSVFCCKEHHG-LTFQ 205  
 QY 383 NIKVLPTWSTPVPCPAEPKSCDKTHTCPBLGGPSVFLPPPKQKTLMTSRPEVTGV 442  
 Db 206 NASMCTSDQPV-----GISFTIPPS-FASIFPTKSAKSLCV 243  
 QY 443 VDVSHEDPEVKFNVYGVGVHNAKTPREQVNSTRVVSVLTVLHODMLNGEYKCY 502  
 Db 244 TDLATYV-STVISTRENGALKTHINISHPNGITSANGELATVCVEEBSGEOFTCV 302  
 QY 503 SNKALPAPIEKTSKAG-QPREPVYTLPPSRDEL-TKNQVSLTCLVKGFPSDIAMV 560  
 Db 303 THDLPVLTQKTSIRPQVAVHMSYVVLPPSREGQLRSATLSCLTVTSPPDVVQ 362  
 QY 561 ESNQO--PENNYKTPVLV--SDGSFLLSKLTVDKSRMQQGVSCSVMEALAHNYT 616  
 Db 363 VQKQGVPPDPSYVTSAMPBPGAGLYFAHSILTVSEEWNAGETVTCVVAHSELPARRVT 422

QY 617 OKSLSPG-----LQDDETCAB 634  
 Db 423 ERSVDSKCKPVLVNVSLVSLDTAG 448

## RESULT 77

B30503  
 Ig gamma-2a chain C region (B5.7A12) - mouse (fragment)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 31-Mar-1989 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999  
 C/Accession: B30503  
 R/Gilmore, G.L.; Bard, J.A.; Birshstein, B.K.  
 J. Immunol. 141, 1754-1761, 1988  
 A/Title: DNA rearrangements affecting both variable and constant regions of Ig H chain.  
 A/Reference number: A30503; MUID:88315788; PMID:2842402  
 A/Accession: B30503  
 A/Molecule type: mRNA  
 A/Residues: 1-112 <GIL>  
 A/Cross-references: GB:M21925  
 A/Experimental source: myeloma cell line MPC11  
 A/Note: the authors translated the codon GAG for residue residue 11 as Ser  
 C/Genetics:  
 A/Introns: 100/3  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin  
 F/1-70/Domain: immunoglobulin homology <IMM>

Query Match 9.7%; Score 331.5; DB 2; Length 112;

Best Local Similarity 50.0%; Pred. No. 5,2e-13;

Matches 66; Conservative 17; Mismatches 22; Indels 27; Gaps 2;

QY 433 SRTPEVTCVNVVSHDPEKFNWYVDGVEVNAKTKPREEQYNSTRVSVTVLHODM 492  
 Db 1 SLTPKTCVNVVSEDDPVQLSWFNVNVETTAQQTIRREDYNSTRVSVTVLHODM 60  
 QY 493 LNKGEYKCKVSNKALPAPIEKTSKAGQPREPVYTLPPSRDELTKQVSLTCLVKGFY 552  
 Db 61 MGKKEKCKVNNKDLPAPIERTISKPGK-----SCSLIAMLGLW- 100.  
 QY 553 PSDIAMEBSNG 564  
 Db 101 -----WTSNG 105

## RESULT 78

S60266  
 novel antigen receptor precursor - nurse shark  
 C/Species: Ginglymostoma cirratum (nurse shark)  
 C/Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 21-Jul-2000  
 C/Accession: S60266  
 R/Greenberg, A.S.; Avila, D.; Hughes, M.; Hughes, A.; McKimney, E.C.; Flajnik, M.F.  
 Nature 378, 168-173, 1995  
 A/Title: A new antigen receptor gene family that undergoes rearrangement and extensive  
 A/Reference number: S60266; MUID:95183140; PMID:7877689  
 A/Accession: S60266  
 A/Status: preliminary; nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-684 <GRE>  
 A/Cross-references: EMBL:U18701; NID:g699442; PIDN:AA848195.1; PID:g699443

Query Match 9.6%; Score 326.5; DB 2; Length 684;

Best Local Similarity 23.8%; Pred. No. 1e-11;

Matches 156; Conservative 86; Mismatches 253; Indels 161; Gaps 29;

QY 16 LALLPAATQGNKRVILGKKDVTLELTCTAS--QKSIQFMW-KNSNOIKILNGSGFLTKG 72  
 Db 145 VSLHSGATSEQA-----NRFQVLVCLISGYPENIAVAMQKNTKTI-----TSGPATIS 194  
 QY 73 PSKTLNRAASRSLSL-----WDGQFPLIKLKIKEDSDTYICEVEDQKEVQLVFLG 125  
 Db 195 PVTSSNDPSCASLLKVPLEQWMSRGS-----VYSCQV----- 226  
 QY 126 TANSDFHLQGSLLTLLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWCT 185

```

Db 227 -SHGATSSNQKEIRST-----SEIAYLADP-----TVEIWDIKSATLIVCE 268
Qy 186 VLQKQKVEPKIDIVLAFQKASSIVYKGEQVE---FSPFLAFYBKLTGSELWQA 242
Db 269 VLSSTVASG-----VVVSMVNGVY--RNEGQMEPTKMSGNOYLITISRLSSVEEM--- 317
Qy 243 ERASSSKWITFDLKNKVSRYVTDOPKLOMGKKLPLHLTLPOALPOVAGSGNLTLALE 302
Db 318 ----QSGVEYTCSAKODOSTTPVVKYKTRKKAVEREPTK-HLRLPSPBEIOSTS----- 366
Qy 303 AKTGKTLHQEVNLVYMRATOLOKNLTCVEWGPSTSPKMLSLKLENKAKVSKREKRVWLN 362
Db 367 -----SATLTCLIRGFYPRK--VSVMQODVSVSN-----VTN 399
Qy 363 PEAGMOCCLSDSGOVLIESNIKVLPTWST--PYPCPAPEPKS-----CDK 406
Db 400 PPTALBDPLTSTSL---NLTVAV-EWKSQAKYCTASHPPSOSTVYKRVIRNQKVDQRQ 455
Qy 407 THTPELLGSPSVFLPPEPKDPTLMISRTPEVTCVVVDVSHEDPE-VKFNMYVDGVEVHN 465
Db 456 TDI-----SVSLKLP-PPEEITWTOQTATIVCEIV--YSDLENIKVQVNGVERKK 503
Qy 466 AKTRPREQVNSTYRVVSVLTVLHQDMLNGKEYCKVSNKALPAPIEKTISKAK-GQPRE 524
Db 504 GVTQNPPEWMSGSKTIYSKLVNMASEWDSGEYVCLVEDSELPTPVKASIRKANVSGMHP 563
Qy 525 PQVYTLTPPSRDEL-TKNQVSLTCLVKGFYPSDIAVEMESNGQ-PENNYKTPPYLSDGS 582
Db 564 PKVYLLHPSTDEITENSATLMCLATNHPAIIYVGMANADTLDSGVRITQVDESEKSGS 623
Qy 583 FELYSKLTVDKSRNQGVNFSQVMEHALH-----NHYQKSLSPGLQDET 631
Db 624 SFVTDRLKLTIAEWNSTTYSCLVGHPSLNRDLIRSTKNSGKSSVNVSVLSDT 679

```

## RESULT 79

```

C31933
Ig mu chain C region - African clawed frog (fragment)
C/Species: Xenopus laevis (African clawed frog)
C/Date: 31-Mar-1990 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
C/Accession: C31933
R/Schwager, J.; Mikoyak, C.A.; Steiner, L.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 2245-2249, 1988
A/Title: Amino acid sequence of heavy chain from Xenopus laevis IgM deduced from cDNA se
A/Reference number: A94192; MUID:88176921; PMID:2451244
A/Accession: C31933
A/Molecule type: mRNA
A/Residues: 1-453 <SC>
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin

```

```

Query Match 9.5%; Score 324.5; DB 2; Length 453;
Best Local Similarity 29.5%; Pred. No. 8e-12;
Matches 98; Conservative 54; Mismatches 115; Indels 65; Gaps 14;
Qy 342 LKLENKAKVSKREKRV-----WVLNPEAGMOCCL--SDSGOVL 380
Db 146 LKNGNQTTGEGVREEPVDEKRGYEATSYLSITRKEMLDLT--LYSCVVEHABSGS-LQ 201
Qy 381 ESNIKVLPTWSTPYPCPAPEPKSCDKHTTCELLGSPVFLPPEPKDPTLMISRTPEVTC 440
Db 202 EKNM-----SKSLMCDTPITPTSIQVITIP-----PSL-----ESIFEKKSATLTC 242
Qy 441 VVVVDVSHEDPEVKFNMY-VDGEVNAVNAKTKPREQVNS--TYRVSVLTVLHQDMLNGKE 497
Db 243 LVSMANSEDLRSISMFKSGSTQELPLKTEIGDAIYNDRNRYSVKGTITVQADAMNDK- 301
Qy 498 YKCVSNKALPAPIEKTISKAKGQPREQVYTLTPPSRDELTKNQ-VSLTCLVKGFYPSDI 556
Db 302 FVCKVEHETLASMKVEFLFKEKEGYNTPSVVPPELLELSKRETAATLTCLVKGFPSEI 361
Qy 557 AVEM--ESNGQPENNYKTP-----PYLSDSGSFYLSKLTVDKSRNQGVNFSQVMEH 608

```

```

Db 362 FVKMLHNGEAVPRQNYINTSINDELPRKGQSKFPLYSLHTIDIKWDAGDSFSCVGH 421
Qy 609 EALHNHYTQKSLSPG-----LOLDET 632
Db 422 ESLPLQTLORSIDKSSGKPTNVNVSLVSLDTC 453

```

## RESULT 80

```

S31436
Ig epsilon chain - axolotl (fragment)
C/Species: Ambystoma mexicanum (axolotl)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C/Accession: S31436
R/Fellah, U.S.; Miles, M.V.; Schwager, J.; Charlemagne, J.
Submitted to the EMBL Data Library, November 1992
A/Description: cDNA sequence of Ambystoma mexicanum epsilon heavy Ig chain.
A/Reference number: S31436
A/Accession: S31436
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-433 <FE>
A/Cross-references: EMBL:X69492; NID:G62420; PID:G62421
C/Superfamily: immunoglobulin C region; immunoglobulin homology

```

```

Query Match 9.4%; Score 320.5; DB 2; Length 433;
Best Local Similarity 24.2%; Pred. No. 1.3e-11;
Matches 108; Conservative 65; Mismatches 162; Indels 111; Gaps 19;

```

```

Qy 216 GEQVSEFPPLAFYVEKLTSGELMWQAEKSSKSS-----WIT-----FDLKNKE 260
Db 34 GEPEVSWKAGSTT---MPGKTFPRAAQCATLSTSSQIRIRAYEKGNTSYSGSVGHK 90
Qy 261 VSV-----KRVTDPKLOMGKKLPLHLTLPOALPOVAGSGNLTLALEKTKLH 309
Db 91 TSTEIHKTITSACECKATSKPSVQV-----LQSSCADTQNGSI----- 129
Qy 310 QEVNVLVYMRATOLOKNLTCVEWGPSTSPKMLSLKLENKAKVSKREKRVWVLNPEAGMQ 369
Db 130 -----ELVCLISGTPDNIDVRLVNDKMPIDQQTSP-----PQK--- 165
Qy 370 CLVSDGOVLIESNIKVLPT-WST-----PYPCPAPEPKSCDKHTTCELLG--GPSVF 420
Db 166 ---DGGTFSTTQIINTKSDMASGDKYTKGVENHPATSSAEDLTINHCADSQIRPYQKVF 222
Qy 421 LPEPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAVNAKTKPRE---EQY 475
Db 223 LIAPKARD-LYIANQPVICKITKEMNSD-SLSVTW-----KREGPEEAIVISEQY 272
Qy 476 ---NSTRYVSVLTVLHQDMLNGKEYCKVSNKALPAPIEKTISKAKGQPREQVYTLTP 532
Db 273 IDSDGFTTAMSYNITNEMERGDDEFCKVGHDLPEPLRSVSKPTGR3FAFTMYFAP 332
Qy 533 SRDELTK-NOVSLTCLVKGFYPSDIAVEMESNGQ--PENNYKTPPYLSDD-----GSFPL 585
Db 333 HEMELANYDVSLTCLVKSFPDDIYIQMGQGSVLPDSKTVSMERPOEAGTAGLGYFS 392
Qy 586 YSKLTVDKSRNQGVNFSQVMEHAL 611
Db 393 YSMLTICKSDMDKREFPTCVAAHSAV 418

```

## RESULT 81

```

AHRB
Ig alpha chain C region - rabbit (fragment)
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 16-Jul-1999
C/Accession: A02174
R/Knight, K.L.; Martens, C.L.; Stoklosa, C.M.; Schneiderman, R.D.
Nucleic Acids Res. 12, 1657-1670, 1984
A/Title: Genes encoding alpha-heavy chains of rabbit IgA: characterization of cDNA enco
A/Reference number: A02174; MUID:84144059; PMID:6322114
A/Accession: A02174

```

A:Molecule type: mRNA  
A:Residues: 1-299 <NM1>  
A:Cross-references: GB:X00353; NID:g1575; PIDN:CAA25100.1; PID:g1576  
C:Comment: This immunoglobulin belongs to the Iga-g subclass. It was isolated from a rat  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)  
chain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1a  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin; plasma  
F:86-152/Domain: immunoglobulin homology <IM1>  
F:189-261/Domain: immunoglobulin homology <IM2>  
F:38,286/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.0%; Score 306; DB 1; Length 299;  
Best Local Similarity 32.2%; Pred. No. 5,7e-11;  
Matches 88; Conservative 42; Mismatches 115; Indels 28; Gaps 13;

QY 369 QCLISDQGVLLSNIKVLPTWSTPVP--CPAPEKSCDKHTPC-----ELIG--PSV 419  
DB 20 QCLGQKSAACHVEVNSVINESLPVPFDDCPA---NSC---CTCPSSSRNLISGQPSL 73  
QY 420 FLPPPKKDTLMSRTPETCVVVDVSHEDPEVKFNWYDVGVEVNAKTRPEEQYNSTY 479  
DB 74 SLQRPDLGD--LLIGRDASLTCTLSGLKNPEDAV--FTW--EPTNGNEPVQORARDISGCY 129  
QY 480 RVSVVLTVLHQMUNGKEVCKVSNKALPA--PIEKTISKAKQPREPOVYTLPPSRDEL 538  
DB 130 SVSSVLPSSAEWKARTEFTCTVTHPEIDSGSLTATISRGVTP--PQVHLLPPESELA 187  
QY 539 KN-QVSLTCLVGFPSYSDIAVEWESNGQ--PENNY---KTPPVLDSDGSFLYSKLTVD 592  
DB 188 LNEQVTLTCLVGFSPKDVLSVRHQGEVPEDSFLVWKSMPSSQDKATYATISLRVP 247  
QY 593 KSRMOGNYVSCSVNHEALNHYTKSLSLSPG 625  
DB 248 AEDMNGDTYSCWVGHEGLAEHFTQRTIDRLAG 280

RESULT 82  
S21461  
T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 20-Feb-1995 #sequence\_revision 19-Apr-1996 #text\_change 21-Jan-2000  
C/Accession: I47131; S21461  
R:Guetafason, K.; Germana, S.; Sundt, T.M.  
J. Immunol. 151, 1365-1370, 1993  
A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine  
A:Reference number: I47131; MUID:93329116; PMID:8335933  
A:Accession: I47131  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-99 <GU2>  
A:Cross-references: EMBL:X65629; NID:g1928; PIDN:CAA46593.1; PID:g388232  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C:Keywords: glycoprotein; T-cell  
F:3-81/Domain: immunoglobulin homology <IMM>

Query Match 8.9%; Score 305.5; DB 2; Length 99;  
Best Local Similarity 60.2%; Pred. No. 1.5e-11;  
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 32 KKSDTVELCTASQKSIOPFMKNSNQIKILNGSFL--TKGPSKLINDRADSRSLMDQG 90  
DB 1 KAGDIALPCHSSQKSLPFNMKNSNQIKILGHSWMHASTELTSLRDSKKMWDHG 60

QY 91 NPELLIKLKIEDSDYICEVEDQKEVQLLYVGLTAN 128  
DB 61 SFLIIKMLEVTDGIVCEVEDKRIEVLQVLRFTAS 98

RESULT 83  
S09276  
Ig alpha chain C region - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
C/Accession: S09276  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A:Reference number: S09264; MUID:90076124; PMID:2512120  
A:Accession: S09276  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-338 <BUR>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:228-300/Domain: immunoglobulin homology <IMM>

Query Match 8.9%; Score 303; DB 2; Length 338;  
Best Local Similarity 31.9%; Pred. No. 1e-10;  
Matches 87; Conservative 43; Mismatches 115; Indels 28; Gaps 13;

QY 369 QCLISDQGVLLSNIKVLPTWSTPVP--CPAPEKSCDKHTPC-----ELIG--PSV 419  
DB 59 QCLGQKSAACHVEVNSVINESLPVPFDDCPA---NSC---CTCPSSSRNLISGQPSL 112  
QY 420 FLPPPKKDTLMSRTPETCVVVDVSHEDPEVKFNWYDVGVEVNAKTRPEEQYNSTY 479  
DB 113 SLQRPDLGD--LLIGRDASLTCTLSGLKNPEDAV--FTW--EPTNGNEPVQORARDISGCY 168  
QY 480 RVSVVLTVLHQMUNGKEVCKVSNKALPA--PIEKTISKAKQPREPOVYTLPPSRDEL 538  
DB 169 SVSSVLPSSAEWKARTEFTCTVTHPEIDSGSLTATISRGVTP--PQVHLLPPESELA 226  
QY 539 KN-QVSLTCLVGFPSYSDIAVEWESNGQ--PENNY---KTPPVLDSDGSFLYSKLTVD 592  
DB 227 LNEQVTLTCLVGFSPKDVLSVRHQGEVPEDSFLVWKSMPSSQDKATYATISLRVP 286  
QY 593 KSRMOGNYVSCSVNHEALNHYTKSLSLSPG 625  
DB 287 AEDMNGDTYSCWVGHEGLAEHFTQRTIDRLAG 319

RESULT 84  
HVKKCO  
Ig mu chain C region, membrane-bound (clone 3050) - horn shark  
C:Species: Heterodontus francisci (horn shark)  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Aug-1996  
C/Accession: S01854; C32716; A46530  
R:Kokubu, F.; Hinder, K.; Litman, R.; Shambloot, M.J.; Litman, G.W.  
EMBO J. 7, 1979-1988, 1988  
A:Title: Complete structure and organization of immunoglobulin heavy chain constant region  
A:Reference number: S00980; MUID:88328985; PMID:3138109  
A:Accession: S01854  
A:Molecule type: DNA  
A:Residues: 1-461 <KOK>  
A:Cross-references: EMBL:X07781  
A:Note: The sequence was determined from the germline gene  
R:Kokubu, F.; Hinder, K.; Litman, R.; Shambloot, M.J.; Litman, G.W.  
Proc. Natl. Acad. Sci. U.S.A. 84, 5868-5872, 1987  
A:Title: Extensive families of constant region genes in a phylogenetically primitive vertebrate  
A:Reference number: A32716; MUID:87289703; PMID:3475706  
A:Accession: C32716  
A:Molecule type: DNA  
A:Residues: 1-99 <KO2>  
A:Cross-references: GB:M17186  
C:Genetics: 100/1; 206/1; 309/1; 419/1; 459/3  
A:Introns: 100/1; 206/1; 309/1; 419/1; 459/3  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)  
chain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1a  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin  
F:20-87/Domain: immunoglobulin homology <IMM1>  
F:123-190/Domain: immunoglobulin homology <IMM2>  
F:228-291/Domain: immunoglobulin homology <IMM3>  
F:330-400/Domain: immunoglobulin homology <IMM4>  
F:438-458/Domain: transmembrane #status predicted <TMW>





```

Db      219 LGNN-EPVCAAGAAATCA-----DVKETTGGDVCVCIPIFSKVTLLSLDPPQEDERRVLLV 272
Qy      283 TLPOALPOVAGSGNLTALAEAKTGKLGHEVNLYVMRATQIQKRLTCEWGPISPKMLSL 342
Db      273 CLVEGLPFS-AGA-----AIQ-----W----- 287
Qy      343 KLENKAKYKSKREKPVVNLNPEAGMOCLLSDGQVLLBSNIVL-PTW----- 390
Db      288 -LQDNBEMTPAPE-----SDSGCSDC--TESG-VTQMSRVVTKRSGEGAGFCGRVT 337
Qy      391 --STVPVCPAPEKSCDKHTTCPELLGSPVFLFPKPKDTLMISRTPEVTVVVDVSH 448
Db      338 HGLKKEPVTVATVETDCAAT--POL-----QVSLPPTLEB-LIVSNATVTCVSNAAAA 389
Qy      449 DPEVKENWYVD---GVEVHNATKPREEQYNSTYRVSVLTVLHQMLNKEYKCKVSNK 505
Db      390 D-GVSVSWSRSSGGLDV-----SQTEDRQADGRYTVRSFLRVCAGEENNGETFGCSVRE 444
Qy      506 ALPAPLEKTSKAGQP-REPOVYTLPPSRDELTKNOVSLTGLVGFYPSDIAVEMSN 563
Db      445 GV-VVAEESIRKKTDPPLHAPSVYVPPPAEELSLQETATVLTCMASFLPSSILLTWTOQ 503
Qy      564 GQP--ENNYKTTPPVLDSDGSFF-LYSKLTVDKSRMOQGNVFCSVMEALHNHYTKSL 620
Db      504 NQDISPQNVILFGR--EKQDGFYSYKSLKLVSVEDMQRGDFGCVGVGHDSIPLNFIHKSI 561
Qy      621 SLSPG 625
Db      562 DKNAQ 566

```

## RESULT 87

```

509270
Ig alpha chain C region - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 29-Jan-1993 #sequence_rev150 29-Jan-1993 #text_change 16-Jul-1999
C:Accession: S09270
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
EMBO J. 8, 4041-4047, 1989
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13
A:Reference number: S09264; MUID:90076124; PMID:2512120
A:Accession: S09270
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-348 <BOR>
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:133-199/Domain: immunoglobulin homology <IM>

```

```

Query Match      8.7%; Score 298.5; DB 2; Length 348;
Best Local Similarity 32.6%; Pred. No. 1.9e-10;
Matches 87; Conservative 41; Mismatches 114; Indels 25; Gaps 13;

Qy      369 QCLLSDSGQVLLBSNIVLPTWSTPVPVCPAPEKSCDKHT--TCPELLGSPVFLFPKPK 426
Db      78 QCLEYDPAACHVEYNVI--NESLPVPFPDP---CEQCHCPSCGE---PSLSLQRPDL 127
Qy      427 KDTLMISRTPEVTVVVDVSHEDPEVKENY-VGVEVHNATKPREEQYNSTYRVSVL 485
Db      128 RD-LLLGSDASLFTTLRGLKYPEDAV-FWEPTNGNEF--VQSPQRDPC-GCYVSSVL 182
Qy      486 TVLHQMLNKEYKCKVSNKALP-APLEKTSKAGQPREPOVYTLPPSRDELTKNQ-VS 543
Db      183 PGCAEPNNACTEFTCTVTHPEIBGSLTATISKDTGSLTPPVHLLPPESEELALNALVT 242
Qy      544 LTCLVKGFPYSDIAVEMSNQ--OPENNY--KTTPPVLDSDGSFFLYSLGLTVDKSRMOQ 598
Db      243 LTCLVKGFPYSDIAVEMSNQ--OPENNY--KTTPPVLDSDGSFFLYSLGLTVDKSRMOQ 598
Qy      599 GNVFSSGVMEALHNHYTKSLSPG 625
Db      303 GDSYTCVVGHEGLAEHFTQRTIDREAG 329

```

## RESULT 88

```

147175
Ig alpha chain C region - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 21-Feb-1997 #sequence_rev150 21-Feb-1997 #text_change 21-Jan-2000
C:Accession: I47175
R:Brown, W.R.; Butler, J.E.
Mol. Immunol. 31, 633-642, 1994
A:Title: Characterization of a C alpha gene of swine.
A:Reference number: I47175; MUID:94254897; PMID:7545929
A:Accession: I47175
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-342 <BOR>
A:Cross-References: EMBL:U12594; NID:9555826; PIDN:AAA65943.1; PID:9555827
C:Genetics:
A:Gene: IGACalpha
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:232-304/Domain: immunoglobulin homology <IM>

```

```

Query Match      8.7%; Score 298; DB 2; Length 342;
Best Local Similarity 31.8%; Pred. No. 2e-10;
Matches 88; Conservative 42; Mismatches 109; Indels 38; Gaps 13;

Qy      372 LSDSGQVLLBSNI--KVLPTWSTPVPVCPAPEKSCDKHTTCPELLGSPVFLFPKPKDT 429
Db      88 LSKSSQSV---NVPCVLPFS---DPCP-----QCCK-----PSLSLQRPALAD- 124
Qy      430 LMSRTPEVTVVVDVSHEDPEVKENYVDGVEVHNATKPREEQYNSTYRVSVLTVLH 489
Db      125 LLLGSDASLFTTLRGLKYPEDAV-FWEPTNGNEF--VQSPQRDPC-GCYVSSVL 182
Qy      490 QDMLNKEYKCKVSNKALPAPLEKTSKAGQPREPOVYTLPPSRDELTKNQ-VSLTCLV 548
Db      182 DPMNGEFTFSCVAASSELKSLATATITKPKVTFPPVHLLPPESEELALNELVTLCLV 241
Qy      549 KGFYPSDIAVEMSNQ--PENNY--KTTPPVLDSDGSFFLYSLGLTVVDKSRMOQGNVS 603
Db      242 RGFSPDIAVEMSNQ--PENNY--KTTPPVLDSDGSFFLYSLGLTVVDKSRMOQGNVS 603
Qy      604 CSVMEALHNHYTKSLSPG----LQDDECAEAQ 636
Db      302 CMVGHFALPLATQKTRIDRLAKGPTHVNVSVMAEAE 338

```

## RESULT 89

## HVRK2

```

Ig mu chain C region (clone 12022) - horn shark (fragment)
C:Species: Heterodontus francisci (horn shark)
C>Date: 30-Jun-1991 #sequence_rev150 30-Jun-1991 #text_change 16-Jul-1999
C:Accession: S00980
R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloot, M.J.; Litman, G.W.
EMBO J. 7, 1979-1988, 1988
A:Title: Complete structure and organization of immunoglobulin heavy chain constant region
A:Reference number: S00980; MUID:86328985; PMID:3138109
A:Accession: S00980
A:Molecule type: mRNA
A:Residues: 1-438 <KOK>
A:Cross-References: EMBL:X07784; NID:963963; PIDN:CAA30617.1; PID:963964
A:Note: The sequence was determined from the differentially expressed gene
C:Complex: An immunoglobulin heterotrimeric subunit consists of two identical light (L)
chain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into 1-
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
F:1-438/Domain: C region <CR>
F:20-87/Domain: immunoglobulin homology <IM>
F:223-190/Domain: immunoglobulin homology <IM>
F:228-291/Domain: immunoglobulin homology <IM>
F:330-400/Domain: immunoglobulin homology <IM>
F:166,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status p.

```



Query Match	8.7%	Score 297.5;	DB 1;	Length 438;
Best Local Similarity	25.2%	Pred. No. 3e-10;		
Matches 113; Conservative	72;	Mismatches 167;	Indels 97;	Gaps 19

```

Qy 208 SSIYKKEGEQVEFSPLAFY-----EKLTSGEMLWMQERASSKSM 251
Db 37 TSIIWKKDKEPIITGLKIPSVLANKKGYTRSSQULTTESEVSSKITCEVNRGES -LW 94
Qy 252 ITFDLKNKEVSVKRYTODPKLQMGKKLPLHLPLPOLFOYAGSGMLTALAEKTKLHOE 311
Db 95 I-----KEI-----LDCK---GDIVEPTVILITSSSEELTSRRFATVLCIIIDFHPS 139
Qy 312 VNLVYMRATQ-----LQKULTCEWGPSPSPKMLSLKLENKKEAKSKKEKPWVLNPEAG 366
Db 140 ITVSWLKXGQOPMDSGFVTSPTCEVNGNSATSRLTVPAGE-----WFNSNT--- 184
Qy 367 MMOCILSDSGQVILTESNIKVLPTWSPCEPAPKEKCDKHTHCPELLGSPSVFLFPPKP 426
Db 185 VYTCQVANAQ-EYTGSRNIT-----GSQVPCS-----IGDPVITKLLPFSI 222
Qy 427 KDTLMISTRPEYTCVAVDVSHEDPEVKFNWYVDGVEVHNAKTKPREBOY---NSTYRV 481
Db 223 EQVL-LEATVTLITCY---VSNAPYGVANSM-----TOEKKPLKSEIAVQPGEDSDSV 270
Qy 482 VSVLTVLHQDMLNGEKYCKVSNKRLPAPIEKTISKAGOP-REPOV-YTLPBPSDELTK 539
Db 271 ISTVNIISTQAMLSGAEFYCVVSHOULPPLRASIHKEEKVDUREFVSVLLPBPADVSAQ 330
Qy 540 NOVSLTCLVKGFPSPDIAMWESNQGOPEN--NYKTTTPVLDSDG-SFLLYSKLTYDKSRW 596
Db 331 RFLSLTCLVNRGSPREIFIKMTVNDKSNVNGRYKOTBEVAERDNRSFFIYSLLSIAEBW 390
Qy 597 QQGAVFSGSWHEALHNHYTOKSLSLSPG 625
Db 391 ASGASYSVCVNGHEALPLKIINFTVKKSSG 419

```

**RESULT 90**

S09269 IG alpha chain C region - rabbit (fragment)  
C/Species: Oryctolagus cuniculus (domestic rabbit)  
C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #next\_change 16-Jul-1999  
C/Accession: S09269  
R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A/Reference number: S09264; MUID:90076124; PMID:2512120  
A/Accession: S09269  
A/Status: not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-357 <BUR>  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: immunoglobulin  
E/142-208/Domain: immunoglobulin homology <IMM>

Query Match	8.7%	Score 297	DB 2	Length 357
Best Local Similarity	32.9%	Pred. No. 2,4e-10		
Matches	82	Conservative 38	Mismatches 99	Indels 30
			Gaps	11
QY	388	PTWSPVVCAPABEPKSCDKHTCPBLGSPVFLFPPPKRQDTLMISRPBETVCVVVDVSH	447	
Db	109	PSDITTCPCPCSP-SC-----BEPSSLDRPPLRD-LILNNAISLCTTLRGKN	156	
QY	448	EDPEVKFMVYDGVENHAKTRPREEQVNS----TYRVSVLTVLHODMLNGEKYCKRVS	503	
Db	157	PEGAV-FPM-----EPTNG-NKPVQGSVOVSQCGCYSSVLPGCAEPMNNAETFTCTYT	209	
QY	504	NKALP-APLEKTSIAKQGREPOVYVTLPPSRDELTKNQ-VSLTCLVGFVPSDIAVME	561	
Db	210	HPEITGSGVLTAKISDQTGAIIPQVHLPLPPEBELALNELVTLTLCVGFSGKDVLVWT	269	
QY	562	SNG--QPPNNY--KTPPVLVDSGSPFLYSKLTIVDKSRMOQGNVSGSVNHEALHNHYT	616	

Db 270 NKGWNPENSLVMKPLPBPGEPTTAVTSLLRVPAEDNNQNESTYCVGHESLAHFT 329

QY 617 QKSLSLSPG 625

Db 330 QRTIDRLSG 338

## RESULT 91

S09274  
IG alpha chain C region - rabbit (Fragment)  
C/Species: Oryctolagus cuniculus (domestic rabbit)  
C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
C/Accession: S09274  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A/Title: The IGA heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A/Reference number: S09264; MUID:90076124; PMID:2512120  
A/Accession: S09274  
A/Status: not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-347 <BUR>  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: immunoglobulin  
P/132-198/Domain: immunoglobulin homology <IMM>

Query Match	8.7%	Score 295.5	DB 2	Length 347
Best Local Similarity	33.2%	Pred. No. 2.9e-10		
Matches	89	Conservative	40	Mismatches 112; Indels 27; Gaps 14
Qy	369	QCLTSDSQVVLSENIKVLPTWSTPVPCPAPEPKSCDKTH--TCPELLGGSVFLFPPKP	426	
Db	77	QCLEYDSAAACHVEINSVI--NESLPVPPDP---CEGCHGPSCE---PSLSIQRPDL	126	
Qy	427	KDTLMISTREPVTCVVVDVSHEDPE-VKFN-N-YVDGVEVHNAKTRPREEQYNSTYRVVSV	484	
Db	127	RD-LLLGSDASLTTLNGL--KDEGAVFTMGPIINGN-N-PVQGSQRDP-C-GCYSVSSV	180	
Qy	485	LTVLHQDLWNGKEYCKKVKVSKALP-APLEKTIKISRAKGQPREQYVTLPPSDELTKNO-V	542	
Db	181	LPGCAEPMNAGTEFTCTVTHPEIGSSSLTATISKDTGSLTPPLVHLLPPPEELNALNLT	240	
Qy	543	SLTLVNGFGFYSDILAWEWSNG--QPEKNY---KTPPVLDSDGSFYLKLTLYDKSRNQ	597	
Db	241	TLTLVNGFGFSKPDVLVMTNKGAVKPKNSFLWKRPLPEPGDPTTYAVTSLLRPAEDWN	300	
Qy	598	QGNVSCSVNHEALHNHYTKSLSLSPG	625	
Db	301	QNESYSCVAHAEGLAEHTQRTIIRLAG	328	

## RESULT 92

B223360  
 Ig alpha-2 chain C region (allotype A2m(1)) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
 C:Accession: B22360  
 R:Planagan, J.G.; Lefranc, M.P.; Rabbits, T.H.  
 Cell 36, 681-688, 1984  
 A:Title: Mechanisms of divergence and convergence of the human immunoglobulin alpha-1  
 A:Reference number: A94653; MUID:84150179; PMID:6421489  
 A:Accession: B22360  
 A:Molecule type: DNA  
 A:Residues: 1-340 <FLA>  
 C:Genetics:  
 A:Gene: GDB:IGHA2  
 A:Cross-references: GDB:119333; OMIM:147000  
 A:Map position: 14q32.33-14q32.33  
 A:Introns: 1/1 103/1; 210/1  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: immunoglobulin  
 /230-302/Domain: immunoglobulin homology <IM>

Query Match 8.6%; Score 293.5; DB 2; Length 340;

Q. E30. GODDESS/VEREIN BARBEREITUNG - VIEL FREIWECHSELN ZWISCHEN DEN VEREINEN E30

Db 361 PEKINRTVVKSSG 374  
::: |

RESULT 95  
S09266  
Ig alpha chain C region - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999

R:Accession: S09266  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A:Reference number: S09266; MUID:90076124; PMID:2512120  
A:Accession: S09266  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-352 <BUR>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:241-314/Domain: immunoglobulin homology <IMM>

Query Match 8.4%; Score 285.5; DB 2; Length 352;  
Best Local Similarity 33.1%; Pred. No. 1.1e-09;  
Matches 89; Conservative 36; Mismatches 101; Indels 43; Gaps 16;

374 DSGVLLSNIKYL-----PTWSTPVPCEPAPEKSCDKTHCTCPPELLGSPVFLPPPK 427  
91 DEGG-----NLTVLPECKDPN-SDPTPCPCP-PIYC-----GEPSTLGRPDIG 133  
428 DPLMSTRPEVCCVVDVSHDEP-VKFNWY-VDGVEVNAKTKPREQYN-STYVVSV 484  
134 D-LLESNASLCTTISGL--KDEGAVFTNPNFTNGEFLVQGST---QSPGCGYSSVS 186  
485 LTVLHQMWLNGEKYCKVSNKALP-APIEKTISKAKGQPREQVYTLPSRDELTKNQ-V 542  
187 LPGCAEPNNAAGREFCTVTHPEIEGSSLATISRSGLT--FQVHLLPPTBELNALV 244  
543 SLTCLVKGFPSPDIAVESNQ--PENNY--KTPPEVLSDSGSFLLYSKLTVDKRW 596  
245 TLTCVLRGSPKDVAVSWTHNGTLVVPKDSFLVWPKPLPEPGQEPPTVAVTSLRVABDM 304  
597 QCGNVSCSVMEALHNHTCKSLSPG 625  
305 NQGDYSYSCVGHGELAHEFTORTIDRLAG 333

RESULT 96  
S09265

Ig alpha chain C region - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
C:Accession: S09265  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A:Reference number: S09264; MUID:90076124; PMID:2512120  
A:Accession: S09265  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-357 <BUR>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:246-319/Domain: immunoglobulin homology <IMM>

Query Match 8.4%; Score 285.5; DB 2; Length 357;  
Best Local Similarity 32.4%; Pred. No. 1.2e-09;  
Matches 83; Conservative 38; Mismatches 104; Indels 31; Gaps 10;

381 ESNIVALTWSTPVCAPAPPEKSCDKTHCTCPPELLGSPVFLPPPKDTMTSRTPEVTC 440  
103 ESTIEPTTPPCPCPCP---SC-----GKPSLIGRPDLGD-LILNSASLTC 148  
441 VVVDVSHDEPVEKFNWYVDGVEVNAKTKPREE--QYNSTYRVVSVLTVLHQMWLNGKE 497  
149 TLRLGLNPEGAV-FTW-----EPTFGKEVQOOSPOLDHCGCYSSVSFLPGCAVLNAGTE 202  
498 YKCKVSNKALP-APIEKTISKAKGQPREQVYTLPSRDELTKNQ-VSLTCLVKGFPYD 555  
203 FTCTVTHPEIEGSSLATISRSGLTIPQVHLLPPTBELNALVTLTCLVLRGSPKD 262

[illegible]

[illegible]

```

RESULT 99
SI2328
I9 heavy chain C region (clone 5301) - horn shark (fragment)
C|Species: Heterodontus francisci (horn shark)
C|Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 21-Jan-2000
C|Accession: SI2328; S01855
R|Litman, R.
submitted to the EMBL Data Library, May 1988
A|Reference number: SI2328
A|Accession: SI2328
A|Molecule type: mRNA
A|Residues: 1-244 <Lit>
A|Cross-references: EMBL:X07785; NID:G63965; PIDN:CAA30618.1; PID:G63966
A|Note: This sequence was determined from the differentiated gene
R|Okubo, J., Hinds, K., Litman, R., Shandloct, M.J., Litman, G.W.
EMBO J. 7, 1979-1988, 1988
A|Title: Complete structure and organization of immunoglobulin heavy chain constant region
A|Reference number: S00980; MUID:88328985; PMID:318109
A|Accession: S01855
A|Molecule type: mRNA
A|Residues: 200-244 <KOK>
A|Cross-references: EMBL:X07785
A|Note: This sequence was determined from the differentiated gene
C|Superfamily: Immunoglobulin C region; immunoglobulin homology
C|Keywords: glycoprotein; heterotetramer; immunoglobulin; transmembrane protein
F.1-244/Domain: C region (fragment) <CRE>
F.111-181/Domain: immunoglobulin homology <IMW>
F.215-241/Domain: transmembrane #status predicted <TMM>
F.26,155,192,196/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 8.3%; Score 284; DB 2; Length 244;
Best Local Similarity 32.2%; Pred. No. 8.7e-10;
Matches 75; Conservative 39; Mismatches 89; Indels 30; Gaps 9;
Qy 428 DTLMSRPEETCVVVDVSHDEPEYKFMVYDVGVENNAKTKPREEQ-----NSTRTVY 482
Db 4 EQVLLEATVTLTCV--VSNAPYGVNVSW-----TQEQKPLKSEITAVQGEDSDSVI 52
Qy 483 SVLTIVLHDDIMNGEKYEKKCVSNKALPAIEKTIKAKQAP-REFQV-YTLPPSRDELTKN 540
Db 53 STVDISASQSWLSGVDFCVVSHQDLPFLRPFHKEKNKKDLREDSVSTLPLPADVGAQR 112
Qy 541 QVSLTCLVKGKGYPSDIAVEMESNGQFEN-NYKTTTPVYVLDSD-GSFLYISKLTLYDKSRWQ 597

```

```

Db      113 FLSLTCLVGRGSPREIFPKMTVYNDKSNVPGNYKNTKEWAAEANDNSYFIYSLISIAAEWA 172
Oy      598 QGNVFSCSVMHEALH---NHYYTKSLSLSPGLDFTCAEADQELDGLWTT 646
Db      173 SGASVSCVGHGHAIPLKIKINRTVANKSSDSS-----DHIIWEDNEEGSNIIWTT 220

RESULT 100
S05500
Ig alpha-1 chain C region - gorilla (fragment)
C:Species: Gorilla gorilla (gorilla)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 23-Jul-1999
C:Accession: S05500
R:Kawamura, S.; Omoto, K.; Ueda, S.
Nucleic Acids Res. 17, 6732, 1989
A:Title: Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene.
A:Reference number: S05500, MUID:89386006; PMID:2506527
A:Accession: S05500
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-352 <KAW>
A:Cross-references: EMBL:X15045; NID:G22900; PIDN:CAA33147.1; PID:G22901
C:Genetics: 102/1; 222/1
A:Introns: 102/1; 222/1
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:242-314/Domain: immunoglobulin homology <IMM>

```

```

Query Match      8.3%; Score 282.5; DB 2; Length 352;
Best Local Similarity 26.5%; Pred. 0.1.7e-09;
Matches 92; Conservative 46; Mismatches 136; Indels 73; Gaps 12;

QY      305  TGIKHQEVNLVVMATQIQ-KNLTCEVWGFTSPKLMISLKENKEAKVSKREKRWVVLN 362
DB      59  SGDIYTTSSQTLTPATOCPPDGKSVTGVNHNHTND----- 92
QY      363  PEAGMGCILSDSQVLLIESINIKYLPHTWSTPVPAPAPEPKSCDKHTTCPELGGSPVELF 422
DB      93  -----SDQVTPPCRVPSTPTPTSPSTPTSPPCCH-----PRSLH 129
QY      423  PPKEDTLMTISRTPEVLCVVVDVSHEDBEVFNMYVDVGVEVNAKTKPREEOYNSTYRVV 482
DB      130  RPALED-LLGSEANLITCTLGL-RDASGVFTFTPSGK--SAVEGPRENDLGGCVSVS 185
QY      483  SVLTVLHODMLNGEYKCYKCKVSNKALPAPIETKISKAGQPREFOYTLPPSRDELTKNQ- 541
DB      186  SVLPGCAEPNMGHTFTCTAAYPPSKPTPLTATLKS--GNMFRPVMHLPPSEBELANL 244
QY      542  VSLFCVWKGFPSPDIAYEWESNGQ--PENNYKTPPYLD--SDG--SFLYSEKLTIVYDSRW 596
DB      245  VTLTCLARGSPKVLVRYMLOGSELEPREKYLTVASROEBSQGTTFTRAVTSLILVAADW 304
QY      597  QQGAVFSCSVNHEALHNHYTQKSLSPG-----LQLDETC 632
DB      305  KKGDPTSCMVGHBALPLAFTOKTIDRLAKPETHVNVSVMAEVDGTC 351

Search completed: August 3, 2004, 13:15:14
Job time : 21.8924 secs

```

Search completed: August 3, 2004, 13:15:14  
Job time : 21.8924 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using bw model

Run on: August 3, 2004, 13:14:00 ; Search time 54.8475 Seconds  
(without alignments)  
3706.029 Million cell updates/sec

Title: SE07  
Perfect score: 3414  
Sequence: 1 MNRGVPRHLLVLQLALP.....DETCAEADGELDGLMTTDP 648

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

## Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2116	62.0	530	8 US-08-485-163-5	Sequence 5, Appl1
2	2116	62.0	530	9 US-09-766-995-4	Sequence 4, Appl1
3	2059	60.3	432	8 US-08-485-163-3	Sequence 3, Appl1
4	2059	60.3	432	9 US-09-766-995-2	Sequence 2, Appl1
5	2036.5	59.7	532	10 US-09-939-537-6	Sequence 6, Appl1
6	2036.5	59.7	532	11 US-09-243-008-6	Sequence 6, Appl1
7	2036.5	59.7	575	10 US-09-939-537-4	Sequence 4, Appl1
8	2036.5	59.7	575	11 US-09-243-008-4	Sequence 4, Appl1
9	2032.5	59.5	462	10 US-09-939-537-5	Sequence 5, Appl1
10	2032.5	59.5	462	11 US-09-243-008-5	Sequence 5, Appl1
11	2029	59.4	398	10 US-09-939-537-29	Sequence 29, Appl1
12	2021	59.2	457	11 US-09-891-119A-9	Sequence 9, Appl1
13	2017	59.1	402	14 US-10-097-044A-1	Sequence 1, Appl1
14	2015	59.0	458	14 US-10-103-597A-39	Sequence 39, Appl1
15	2015	59.0	458	14 US-10-188-444-39	Sequence 39, Appl1

16	2015	59.0	458	14 US-10-207-655-170	Sequence 170, App
17	2007	58.8	458	12 US-10-151-274-3	Sequence 3, Appl1
18	2001	58.6	397	11 US-09-891-119A-2	Sequence 2, Appl1
19	1997	58.5	458	8 US-08-681-219-27	Sequence 27, Appl1
20	1997	58.5	458	11 US-09-230-111C-25	Sequence 25, Appl1
21	1997	58.5	458	14 US-10-092-138-25	Sequence 25, Appl1
22	1904	55.8	434	14 US-10-097-044A-4	Sequence 4, Appl1
23	1891	55.4	370	9 US-09-759-841-6	Sequence 6, Appl1
24	1877.5	55.0	448	14 US-10-024-329-32	Sequence 32, Appl1
25	1338.5	39.2	254	10 US-09-939-537-33	Sequence 33, Appl1
26	1313.5	38.5	617	14 US-10-363-427-18	Sequence 18, Appl1
27	1313.5	38.5	617	14 US-10-363-427-22	Sequence 22, Appl1
28	1275.5	37.4	437	14 US-10-363-427-14	Sequence 14, Appl1
29	1275	37.3	592	9 US-09-935-868-8	Sequence 8, Appl1
30	1275	37.3	592	14 US-10-287-035-8	Sequence 8, Appl1
31	1275	37.3	592	14 US-10-282-162-8	Sequence 8, Appl1
32	1265	37.1	594	9 US-09-815-108-22	Sequence 22, Appl1
33	1265	37.1	594	14 US-10-229-584-22	Sequence 22, Appl1
34	1265	37.1	779	10 US-09-910-600-16	Sequence 16, Appl1
35	1265	37.1	779	10 US-09-910-600-30	Sequence 30, Appl1
36	1260	36.9	504	14 US-10-207-655-348	Sequence 348, App
37	1260	36.9	543	14 US-10-207-655-345	Sequence 345, App
38	1258	36.8	492	14 US-10-207-655-344	Sequence 344, App
39	1254.5	36.7	567	12 US-09-773-877A-12	Sequence 12, Appl1
40	1254.5	36.7	567	12 US-09-773-877A-20	Sequence 20, Appl1
41	1253.5	36.7	4852	12 US-10-412-406-33	Sequence 33, Appl1
42	1252.5	36.7	552	14 US-10-207-655-266	Sequence 266, App
43	1251	36.6	663	14 US-10-412-406-32	Sequence 32, Appl1
44	1251	36.6	768	14 US-10-207-655-350	Sequence 350, App
45	1249	36.6	698	14 US-09-875-338-9	Sequence 9, Appl1
46	1249	36.6	698	14 US-10-077-023-9	Sequence 9, Appl1
47	1245.5	36.5	557	12 US-09-773-877B-14	Sequence 14, Appl1
48	1243	36.4	500	14 US-10-207-655-240	Sequence 240, App
49	1243	36.4	500	14 US-10-207-655-240	Sequence 398, App
50	1243	36.4	622	12 US-10-683-255-2	Sequence 2, Appl1
51	1243	36.4	949	12 US-10-232-838-18	Sequence 18, Appl1
52	1242.5	36.4	480	9 US-09-875-338-5	Sequence 5, Appl1
53	1242.5	36.4	480	14 US-10-077-023-5	Sequence 5, Appl1
54	1242.5	36.4	480	14 US-10-077-023-5	Sequence 15, Appl1
55	1242.5	36.4	499	14 US-10-207-655-148	Sequence 148, App
56	1242.5	36.4	499	14 US-10-053-530-15	Sequence 15, Appl1
57	1242	36.4	634	16 US-10-416-011-2	Sequence 2, Appl1
58	1241.5	36.4	451	9 US-09-875-338-17	Sequence 17, Appl1
59	1241.5	36.4	451	14 US-10-077-023-17	Sequence 17, Appl1
60	1240	36.3	712	16 US-10-679-620-64	Sequence 64, Appl1
61	1239.5	36.3	555	12 US-10-463-260-6	Sequence 6, Appl1
62	1239.5	36.3	631	15 US-10-120-198B-2	Sequence 24, Appl1
63	1239.5	36.3	502	14 US-10-163-427-24	Sequence 8, Appl1
64	1238.5	36.3	465	12 US-10-404-724-8	Sequence 17, Appl1
65	1238	36.3	934	12 US-10-232-838-17	Sequence 17, Appl1
66	1237	36.2	704	14 US-10-357-653-2	Sequence 2, Appl1
67	1237	36.2	915	16 US-10-282-162-52	Sequence 52, Appl1
68	1236	36.2	715	16 US-10-679-620-62	Sequence 62, Appl1
69	1235.5	36.2	497	12 US-10-683-255-6	Sequence 6, Appl1
70	1235.5	36.2	525	14 US-10-683-255-4	Sequence 4, Appl1
71	1234	36.1	500	14 US-10-207-655-397	Sequence 397, App
72	1233	36.1	367	15 US-10-452-646-9	Sequence 9, Appl1
73	1233	36.1	500	14 US-10-207-655-396	Sequence 396, App
74	1233	36.1	502	14 US-10-363-427-20	Sequence 20, Appl1
75	1233	36.1	543	14 US-10-207-655-346	Sequence 346, App
76	1233	36.1	555	12 US-10-107-991B-3	Sequence 3, Appl1
77	1232	36.1	377	14 US-10-363-427-16	Sequence 16, Appl1
78	1231	36.1	600	16 US-10-334-235-38	Sequence 38, Appl1
79	1230.5	36.0	446	12 US-10-435-299-7	Sequence 7, Appl1
80	1230.5	36.0	550	14 US-10-207-655-270	Sequence 270, App
81	1230.5	36.0	972	15 US-10-418-836-38	Sequence 38, Appl1
82	1230.5	36.0	975	15 US-10-418-836-39	Sequence 39, Appl1
83	1229.5	36.0	810	12 US-10-232-838-15	Sequence 15, Appl1
84	1229.5	36.0	949	12 US-10-232-838-15	Sequence 19, Appl1
85	1229	36.0	482	14 US-10-207-655-307	Sequence 107, App
86	1229	36.0	362	14 US-09-773-877A-18	Sequence 18, Appl1
87	1228.5	36.0	272	12 US-10-433-108-29	Sequence 29, Appl1
88	1228.5	36.0	465	12 US-10-404-724-23	Sequence 23, Appl1

```

89 1228 36.0 467 16 US-10-656-769-32.
90 1228 36.0 476 9 US-09-747-669-3
91 1228 36.0 476 14 US-10-290-703-3
92 1227.5 36.0 472 12 US-10-433-108-21
93 1227.5 36.0 448 15 US-10-378-567-2
94 1226 35.9 471 15 US-10-108-260A-4285
95 1225.5 35.9 470 15 US-10-108-260A-4292
96 1225.5 35.9 900 14 US-10-282-162-34
97 1225 35.9 488 12 US-10-683-255-12
98 1224.5 35.9 396 14 US-10-193-616-14
99 1224.5 35.9 541 16 US-10-471-151-32
100 1224.5 35.9 558 16 US-10-471-151-31
101 1224 35.9 500 14 US-10-207-655-274
102 1223.5 35.8 280 12 US-10-433-108-27
103 1223.5 35.8 470 14 US-10-216-484-117
104 1223.5 35.8 470 14 US-10-216-484-113
105 1223.5 35.8 470 14 US-10-384-933-117
106 1223.5 35.8 470 14 US-10-384-933-113
107 1223.5 35.8 499 14 US-10-207-655-17
108 1223.5 35.8 489 14 US-10-053-530-17
109 1223 35.8 465 12 US-10-404-724-25
110 1223 35.8 480 14 US-10-077-023-133
111 1223 35.8 480 14 US-10-077-023-135
112 1222.5 35.8 445 14 US-10-320-231A-79
113 1222.5 35.8 470 14 US-10-216-484-147
114 1222.5 35.8 470 14 US-10-384-933-147
115 1222.5 35.8 475 9 US-09-740-002-25
116 1222.5 35.8 475 16 US-10-325-698-25
117 1222.5 35.8 579 14 US-10-138-727A-41
118 1222.5 35.8 680 8 US-08-469-583A-15
119 1222 35.8 915 14 US-10-282-162-46
120 1221.5 35.8 444 14 US-10-150-475A-6
121 1221.5 35.8 444 16 US-10-704-522-6
122 1221.5 35.8 444 16 US-10-645-215-6
123 1221.5 35.8 470 14 US-10-216-484-145
124 1221.5 35.8 470 14 US-10-384-933-145
125 1221 35.8 474 15 US-10-108-260A-4282

```

## ALIGNMENTS

```

Sequence 32, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 4285, Ap
Sequence 4292, Ap
Sequence 34, Appl
Sequence 12, Appl
Sequence 14, Appl
Sequence 32, Appl
Sequence 274, Appl
Sequence 27, Appl
Sequence 117, App
Sequence 117, App
Sequence 117, App
Sequence 143, App
Sequence 17, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 135, App
Sequence 79, Appl
Sequence 147, App
Sequence 147, App
Sequence 25, Appl
Sequence 25, Appl
Sequence 41, Appl
Sequence 15, Appl
Sequence 46, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 145, App
Sequence 145, App
Sequence 4282, Ap

```

```

RESULT 1
; Sequence 5, Application US/08485163
; Publication No. US2002098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400

```

```

; TELEFAX: (212) 391-0525
;
; TELE:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-485-163-5

```

```

Query Match 62.0%; Score 2116; DB 8; Length 530;
Best Local Similarity 68.9%; Pred. No. 1.7e-141;
Matches 437; Conservative 25; Mismatches 58; Indels 114; Gaps 10;

```

```

1 MKRGVPRHLILVLTALIPATQGNKVYLGKGGTVELTCTASOKKSIOFMKNSNOIK 60
1 MKRGVPRHLILVLTALIPATQGNKVYLGKGGTVELTCTASOKKSIOFMKNSNOIK 60
61 ILGNQSPFLTKGPSKLNDRADRSRLMDQGNPILIKLIKIEDPTTICEVEDQKEVOL 120
61 ILGNQSPFLTKGPSKLNDRADRSRLMDQGNPILIKLIKIEDPTTICEVEDQKEVOL 120
121 LVFGILTANDTLLDQGSITLTLSPRGSBPVOCRSPPKKNIOGGKTLVSQLEIDSG 180
121 LVFGILTANDTLLDQGSITLTLSPRGSBPVOCRSPPKKNIOGGKTLVSQLEIDSG 180
181 TWTCVTLQKQKVEFIDIVLAFOKASIVYKKEGEQVEFSPFLATVEKLTSGELMW 240
181 TWTCVTLQKQKVEFIDIVLAFOKASIVYKKEGEQVEFSPFLATVEKLTSGELMW 240
181 TWTCVTLQKQKVEFIDIVLAFOKASIVYKKEGEQVEFSPFLATVEKLTSGELMW 240
241 QERASSRSKSWITFDLKNKEVSKVYODPKLQMGKPLHLTLPGALPOYAG--SGNL 297
217 -----PCSRSTSESTAALGCLVNDYPPPYVSMNSGALTSGVH 255
298 TLALEAKTKLQAEVNVLMVMTOL-QKULTEVWCPSTPKMLSLKLNKAKSKRRK 356
256 TEPVALQSSGLYSLSVTVTPSSNFGOTQYTCNV-----DHK 292
357 PVVNLNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV-----PCPAPRPSCKDHTHP 411
293 P-----SNTKVDITVERKCCVECPRCAP----- 317
412 ELGGSVFLFPPEKPDITMISRTPEVTCVVVDVSHEDPEVKFMVYVDGVEVNAATKPR 471
318 --VAGPSVFLPPPKPDITMISRTPEVTCVVVDVSHEDPEVQFMVYVDGVEVNAATKPR 375
472 EEQYNSTYRVSVLTALVDMDLNGEKYKCKVSNKALPAPIETKISAKGQPREPQVYTP 531
376 EEQFNSTYRVSVLTALVDMDLNGEKYKCKVSNKALPAPIETKISAKGQPREPQVYTP 435
532 PSRDELTKQVSLTCLVKGFPYSDIAVENSQGPENNYKTTTPVLDSDGSFELYSLTV 591
436 PSREKTKQVSLTCLVKGFPYSDIAVENSQGPENNYKTTTPVLDSDGSFELYSLTV 495
592 DKSRMQGNVVFSCSVMEALHNHYTQKSLSLSPG 625
496 DKSRMQGNVVFSCSVMEALHNHYTQKSLSLSPG 529

```

```

RESULT 2
; Sequence 4, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
; APPLICANT: Graham P. Allaway et al.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCON
; FILE REFERENCE: 2048/41215-CB/JPM/SHS
; CURRENT APPLICATION NUMBER: US/09/766,995

```

/ CURRENT FILING DATE: 2001-01-22  
 / NUMBER OF SEQ ID NOS: 9  
 / SOFTWARE: Patentin version 3.0  
 / SEQ ID NO 4  
 / LENGTH: 530  
 / TYPE: PRT  
 / ORGANISM: homo sapiens  
 / US-09-766-995-4

Query Match 62.0%; Score 2116; DB 9; Length 530;  
 Best Local Similarity 68.9%; Pred. No. 1.7e-141;  
 Matches 437; Conservative 25; Mismatches 58; Indels 114; Gaps 10;

```

Qy 1 MNRGVPFRHLLVQLALPAAATQGNKVLGKGGDVELTCTASQKSIQFHMNSNQIK 60
Db 1 MNRGVPFRHLLVQLALPAAATQGNKVLGKGGDVELTCTASQKSIQFHMNSNQIK 60
Qy 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGSGLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDTHLLQGSGLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVLQNGKVEFKIDIVLAFAKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
Db 181 TWCTVLQNGKVEFKIDIVLAFAKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
Qy 241 QABRASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPJHLTLPOALPOYAG---SGNL 297
Db 241 QABRASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPJHLTLPOALPOYAG---SGNL 297
Qy 298 TLAEATGKLGHOENVLVVMAATQQLQKNTLCEVWGPTSPKMLSLKENKAAYSKREK 356
Db 298 TLAEATGKLGHOENVLVVMAATQQLQKNTLCEVWGPTSPKMLSLKENKAAYSKREK 356
Qy 357 PAVWLINBAGMWQCLSDSGVLLSNIKVLPTWSTPV-----PCPAPRPSCKDTHNCP 411
Db 357 PAVWLINBAGMWQCLSDSGVLLSNIKVLPTWSTPV-----PCPAPRPSCKDTHNCP 411
Qy 412 ELGGSFVFLPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPR 471
Db 412 ELGGSFVFLPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPR 471
Qy 318 --VAGPSVFLPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPR 375
Db 318 --VAGPSVFLPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPR 375
Qy 472 EEOQNSTFRVVSUTLVHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLR 531
Db 472 EEOQNSTFRVVSUTLVHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLR 531
Qy 376 EEOQNSTFRVVSUTLVHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLR 435
Db 376 EEOQNSTFRVVSUTLVHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLR 435
Qy 532 PSRDELTKNOVSLCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTV 591
Db 532 PSRDELTKNOVSLCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTV 591
Qy 436 PSRDELTKNOVSLCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTV 495
Db 436 PSRDELTKNOVSLCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTV 495
Qy 592 DKSRMOQGNVPSCSVMHEALHNHYTOKSLSLSPG 625
Db 592 DKSRMOQGNVPSCSVMHEALHNHYTOKSLSLSPG 625
Qy 496 DKSRMOQGNVPSCSVMHEALHNHYTOKSLSLSPG 529
Db 496 DKSRMOQGNVPSCSVMHEALHNHYTOKSLSLSPG 529

```

RESULT 3  
 US-08-485-163-3  
 / Sequence 3, Application US/08485163  
 / Publication No. US20020098191A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Beaudry, Gary A.  
 / APPLICANT: Maddon, Paul J.  
 / TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS  
 / NUMBER OF SEQUENCES: 10  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Cooper & Dunham LLP  
 / STREET: 1185 Avenue of the Americas  
 / CITY: New York  
 / STATE: New York  
 / COUNTRY: USA

```

/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/485,163
/ FILING DATE: 07-JUN-1995
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 278-0400
/ TELEFAX: (212) 391-0525
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 432 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ ORIGINAL SOURCE:
/ ORGANISM: homo sapien
/ CELL TYPE: lymphocyte
/ US-08-485-163-3

Query Match 60.3%; Score 2059; DB 8; Length 432;
Best Local Similarity 65.6%; Pred. No. 1.4e-137;
Matches 410; Conservative 12; Mismatches 9; Indels 194; Gaps 3;

Qy 1 MNRGVPFRHLLVQLALPAAATQGNKVLGKGGDVELTCTASQKSIQFHMNSNQIK 60
Db 1 MNRGVPFRHLLVQLALPAAATQGNKVLGKGGDVELTCTASQKSIQFHMNSNQIK 60
Qy 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGSGLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDTHLLQGSGLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVLQNGKVEFKIDIVLAFAKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
Db 181 TWCTVLQNGKVEFKIDIVLAFAKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
Qy 241 QABRASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPJHLTLPOALPOYAGSGNLTLA 300
Db 241 QABRASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPJHLTLPOALPOYAGSGNLTLA 300
Qy 211 -----
Db 211 -----
Qy 301 LEAKTGKLGHOENVLVVMAATQQLQKNTLCEVWGPTSPKMLSLKENKAAYSKREKPVW 360
Db 301 LEAKTGKLGHOENVLVVMAATQQLQKNTLCEVWGPTSPKMLSLKENKAAYSKREKPVW 360
Qy 211 -----
Db 211 -----
Qy 361 LNPBAGMWQCLSDSGVLLSNIKVLPTWSTPVPCPAPRPSCKDTHNCPBELLGSPSVF 420
Db 361 LNPBAGMWQCLSDSGVLLSNIKVLPTWSTPVPCPAPRPSCKDTHNCPBELLGSPSVF 420
Qy 211 -----EC-----PCPAPR-----VAGPSVF 226
Db 211 -----EC-----PCPAPR-----VAGPSVF 226
Qy 421 LPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPREQVYTLR 480
Db 421 LPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPREQVYTLR 480
Qy 227 LPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPREQVYTLR 286
Db 227 LPPPKDITLMSRTPEVTCVVDVSHEDPEVKFMYVDGVEVNAKTKPREQVYTLR 286
Qy 481 VVSVLTVLHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLRPSRDELTKN 540
Db 481 VVSVLTVLHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLRPSRDELTKN 540
Qy 287 VVSVLTVLHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLRPSRDELTKN 346
Db 287 VVSVLTVLHODWLNKKEYCKVSNKALPAPIEKTISSKAGQPREQVYTLRPSRDELTKN 346
Qy 541 QVSLTCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTVDSRMOQGN 600
Db 541 QVSLTCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTVDSRMOQGN 600
Qy 347 QVSLTCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTVDSRMOQGN 406
Db 347 QVSLTCLVKGFPSPDIAVEMESNGOPENNYKTPPLDSDGSFFLYSKLTVDSRMOQGN 406

```

QY 601 VFSCSVHHEALHNHYTOKSLSPG 625  
DB 407 VFSCSVHHEALHNHYTOKSLSPG 431

RESULT 4  
US-09-766-995-2  
Sequence 2, Application US/09766995  
Patent No. US20020052481A1  
GENERAL INFORMATION:  
APPLICANT: Graham P. Allaway et al.  
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IGG2 IMMUNOCONJ  
FILE REFERENCE: 2048/41215-CB/JPM/SHS  
CURRENT APPLICATION NUMBER: US/09/766,995  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 432  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-766-995-2

Query Match 60.3%; Score 2059; DB 9; Length 432;  
Best Local Similarity 65.6%; Pred. No. 1,4e-137;  
Matches 410; Conservative 12; Mismatches 9; Indels 194; Gaps 3;

QY 1 MNRGVFPRHLLVLTQALPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHKNSNOIK 60  
DB 1 MNRGVFPRHLLVLTQALPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHKNSNOIK 60  
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKI ESDTYICEVEDQKEEVOL 120  
DB 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKI ESDTYICEVEDQKEEVOL 120  
QY 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLELODSG 180  
DB 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLELODSG 180  
QY 181 TWICTVLOQOKKVEFIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEKLTGSGELMW 240  
DB 181 TWICTVLOQOKKVEFIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEKLTGSGELMW 240  
QY 241 QABRASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQALPOYAGSGMLTTLA 300  
DB 241 QABRASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQALPOYAGSGMLTTLA 300  
QY 301 LEAKTGLHQBVLVVMRATQLOKNLTCCEVWGPTSPKLMLSKLENKEAKVSKREKPVW 360  
DB 301 LEAKTGLHQBVLVVMRATQLOKNLTCCEVWGPTSPKLMLSKLENKEAKVSKREKPVW 360  
QY 211 -----  
DB 211 -----  
QY 361 LNPBAGMOCILSDSGVLLLESNIKVLPTWSTPVPCBAPKPSGCKHTHTCPELLGSPSVF 420  
DB 361 LNPBAGMOCILSDSGVLLLESNIKVLPTWSTPVPCBAPKPSGCKHTHTCPELLGSPSVF 420  
QY 211 -----EC-----PCCAPP-----VAGGSVF 226  
DB 211 -----EC-----PCCAPP-----VAGGSVF 226  
QY 421 LPPPKPOTLMISRTPEVTCVVVDVSHEDPEVKFNMTYVDGVEYHNAKTKRREEQYNSTR 480  
DB 421 LPPPKPOTLMISRTPEVTCVVVDVSHEDPEVKFNMTYVDGVEYHNAKTKRREEQYNSTR 480  
QY 227 LFPKPOTLMISRTPEVTCVVVDVSHEDPEVKFNMTYVDGVEYHNAKTKRREEQYNSTR 286  
DB 227 LFPKPOTLMISRTPEVTCVVVDVSHEDPEVKFNMTYVDGVEYHNAKTKRREEQYNSTR 286  
QY 481 VVSVLTVLHODMNGKEKCKVCKVSNKALPAPLEKTIISKAKGAPREPOVYTLPPSRDELTKN 540  
DB 481 VVSVLTVLHODMNGKEKCKVCKVSNKALPAPLEKTIISKAKGAPREPOVYTLPPSRDELTKN 540  
QY 287 VVSVLTVLHODMNGKEKCKVCKVSNKALPAPLEKTIISKAKGAPREPOVYTLPPSRDELTKN 346  
DB 287 VVSVLTVLHODMNGKEKCKVCKVSNKALPAPLEKTIISKAKGAPREPOVYTLPPSRDELTKN 346  
QY 541 QVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKTPPVLDSDGSFELYSKLTVDKSRMOQGN 600  
DB 541 QVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKTPPVLDSDGSFELYSKLTVDKSRMOQGN 600  
QY 347 QVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKTPPVLDSDGSFELYSKLTVDKSRMOQGN 406  
DB 347 QVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKTPPVLDSDGSFELYSKLTVDKSRMOQGN 406  
QY 601 VFSCSVHHEALHNHYTOKSLSPG 625  
DB 601 VFSCSVHHEALHNHYTOKSLSPG 625  
QY 407 VFSCSVHHEALHNHYTOKSLSPG 431  
DB 407 VFSCSVHHEALHNHYTOKSLSPG 431

RESULT 5  
US-09-939-537-6  
Sequence 6, Application US/09939537  
Publication No. US20030138410A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian  
Banapour, Babak  
Romeo, Charles  
Kolanus, Waldemar

TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED  
CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS

NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Clark & Elding LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,537  
FILING DATE: 24-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/284,391  
FILING DATE: 02-Aug-1994  
APPLICATION NUMBER: 08/195,395  
FILING DATE: 14-FEB-1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: 06-MAR-1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: 07-MAR-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Elding, Karen L.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/247001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-939-537-6

Query Match 59.7%; Score 2036.5; DB 10; Length 532;  
Best Local Similarity 98.3%; Pred. No. 7.3e-136;  
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVFPRHLLVLTQALPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHKNSNOIK 60  
DB 1 MNRGVFPRHLLVLTQALPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHKNSNOIK 60  
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKI ESDTYICEVEDQKEEVOL 120  
DB 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKI ESDTYICEVEDQKEEVOL 120  
QY 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLELODSG 180  
DB 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLELODSG 180  
QY 181 TWICTVLOQOKKVEFIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEKLTGSGELMW 240  
DB 181 TWICTVLOQOKKVEFIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEKLTGSGELMW 240



Db 181 TWCTVLONQKVEKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Qy 241 QAEARASSKSWITPDLKNKEVSVKRVTDPPKLOQNGKULPLHLTLPQALPOYAGSGNLTIA 300  
Db 241 QAEARASSKSWITPDLKNKEVSVKRVTDPPKLOQNGKULPLHLTLPQALPOYAGSGNLTIA 300  
Qy 301 LEATGKGLHOEVNLVVMARATOLQKNTLCEVWGPTSPKMLSLKLENKAKYSKREKPYWV 360  
Db 301 LEATGKGLHOEVNLVVMARATOLQKNTLCEVWGPTSPKMLSLKLENKAKYSKREKPYWV 360  
Qy 361 LNPEAGMWQCLISDSGQVLLSNIKVLPTWSTPYPCAPBPCKSC 404  
Db 361 LNPEAGMWQCLISDSGQVLLSNIKVLPTWSTPYVHA---DPKLC 401

RESULT 6  
US-09-243-008-6  
Sequence 6, Application US/09243008  
Publication No. US20040005334A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian et al.  
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor Chimeras  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243,008  
FILING DATE: 02-Feb-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,176  
FILING DATE: SEPTEMBER 11, 1995  
APPLICATION NUMBER: 08/203,866  
FILING DATE: February 28, 1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: March 6, 1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: March 7, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Karen F. Lech, Ph.D.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/270001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-243-008-6

Query Match 59.7%; Score 2036.5; DB 11; Length 532;  
Best Local Similarity 98.3%; Pred. No. 7.3e-116;  
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

Qy 1 MNRGVFPHLLLVLTQALLPATQGNKVLKGGDVTVELTCTASQKSIQPHMKNQSIK 60  
Db 1 MNRGVFPHLLLVLTQALLPATQGNKVLKGGDVTVELTCTASQKSIQPHMKNQSIK 60

Qy 61 ILGNQGSFLTKGPSKLNDRADSRRLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYOL 120  
Db 61 ILGNQGSFLTKGPSKLNDRADSRRLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYOL 120  
Qy 121 LVFGLTANSDTHLLIQGSLTTLTSSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
Db 121 LVFGLTANSDTHLLIQGSLTTLTSSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
Qy 181 TWCTVLONQKVEKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWCTVLONQKVEKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Qy 241 QAEARASSKSWITPDLKNKEVSVKRVTDPPKLOQNGKULPLHLTLPQALPOYAGSGNLTIA 300  
Db 241 QAEARASSKSWITPDLKNKEVSVKRVTDPPKLOQNGKULPLHLTLPQALPOYAGSGNLTIA 300  
Qy 301 LEATGKGLHOEVNLVVMARATOLQKNTLCEVWGPTSPKMLSLKLENKAKYSKREKPYWV 360  
Db 301 LEATGKGLHOEVNLVVMARATOLQKNTLCEVWGPTSPKMLSLKLENKAKYSKREKPYWV 360  
Qy 361 LNPEAGMWQCLISDSGQVLLSNIKVLPTWSTPYPCAPBPCKSC 404  
Db 361 LNPEAGMWQCLISDSGQVLLSNIKVLPTWSTPYVHA---DPKLC 401

RESULT 7  
US-09-939-537-4  
Sequence 4, Application US/09939537  
Publication No. US20030138410A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian  
Banapour, Babak  
Romeo, Charles  
Kolanus, Waldemar  
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Clark & Elbing LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PaeSTEO for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,537  
FILING DATE: 24-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/284,391  
FILING DATE: 02-AUG-1994  
APPLICATION NUMBER: 08/195,395  
FILING DATE: 14-FEB-1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: 06-MAR-1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: 07-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elbing, Karen L.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/247001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
 US-09-939-537-4

Query Match 59.7%; Score 2036.5; DB 10; Length 575;  
 Best Local Similarity 98.3%; Pred. No. 8.1e-136;  
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPFRHLILVQLALPAAATQGNKVVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 DB 1 MNRGVPFRHLILVQLALPAAATQGNKVVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 QY 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSDTHLLQGSQSLTLTLSPSSPSVQCRSPRGKNIQGGKTLSSVSOLELQDSG 180  
 DB 121 LVFGLTANSDTHLLQGSQSLTLTLSPSSPSVQCRSPRGKNIQGGKTLSSVSOLELQDSG 180  
 QY 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMW 240  
 DB 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMW 240  
 QY 241 QARRASSSKSWITFDLKNKEVSVKRYTQDPKIQMGKKLPLHLTLPOLQYAGSGNLTLA 300  
 DB 241 QARRASSSKSWITFDLKNKEVSVKRYTQDPKIQMGKKLPLHLTLPOLQYAGSGNLTLA 300  
 QY 301 LEAKTGKLEHQBVLVWMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVWY 360  
 DB 301 LEAKTGKLEHQBVLVWMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVWY 360  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVPAPCPAPKSC 404  
 DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVPAPCPAPKSC 404

RESULT 8  
 US-09-243-008-4  
 Sequence 4, Application US/09243008  
 Publication No. US2004005334A1  
 GENERAL INFORMATION:  
 APPLICANT: Seed, Brian et al.  
 TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor Chimeras  
 NUMBER OF SEQUENCES: 40  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM PS/2 Model 502 or 55SX  
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
 SOFTWARE: Wordperfect (Version 5.0)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/243.008  
 FILING DATE: 02-Feb-1999  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/394.176  
 FILING DATE: SEPTEMBER 11, 1995  
 APPLICATION NUMBER: 08/203,866  
 FILING DATE: February 28, 1994  
 APPLICATION NUMBER: 07/847,566  
 FILING DATE: March 6, 1992  
 APPLICATION NUMBER: 07/665,961

FILING DATE: March 7, 1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Karen F. Lech, Ph.D  
 REGISTRATION NUMBER: 35,238  
 REFERENCE/DOCKET NUMBER: 00786/270001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELE: 200154  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 575 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
 US-09-243-008-4

Query Match 59.7%; Score 2036.5; DB 11; Length 575;  
 Best Local Similarity 98.3%; Pred. No. 8.1e-136;  
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPFRHLILVQLALPAAATQGNKVVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 DB 1 MNRGVPFRHLILVQLALPAAATQGNKVVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 QY 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSDTHLLQGSQSLTLTLSPSSPSVQCRSPRGKNIQGGKTLSSVSOLELQDSG 180  
 DB 121 LVFGLTANSDTHLLQGSQSLTLTLSPSSPSVQCRSPRGKNIQGGKTLSSVSOLELQDSG 180  
 QY 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMW 240  
 DB 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMW 240  
 QY 241 QARRASSSKSWITFDLKNKEVSVKRYTQDPKIQMGKKLPLHLTLPOLQYAGSGNLTLA 300  
 DB 241 QARRASSSKSWITFDLKNKEVSVKRYTQDPKIQMGKKLPLHLTLPOLQYAGSGNLTLA 300  
 QY 301 LEAKTGKLEHQBVLVWMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVWY 360  
 DB 301 LEAKTGKLEHQBVLVWMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVWY 360  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVPAPCPAPKSC 404  
 DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVPAPCPAPKSC 404

RESULT 9  
 US-09-939-537-5  
 Sequence 5, Application US/09939537  
 Publication No. US20030138410A1  
 GENERAL INFORMATION:  
 APPLICANT: Seed, Brian  
 Banapour, Babak  
 Romeo, Charles  
 Kolanus, Waldemar  
 TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS  
 NUMBER OF SEQUENCES: 53  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Clark & Elbing LLP  
 STREET: 176 Federal Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,537  
FILING DATE: 24-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/284,391  
FILING DATE: 02-AUG-1994  
APPLICATION NUMBER: 08/195,395  
FILING DATE: 14-FEB-1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: 06-MAR-1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: 07-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elbing, Karen L.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/247001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-939-537-5

Query Match 59.5%; Score 2032.5; DB 10; Length 462;  
Best Local Similarity 98.0%; Pred. No. 1.2e-135;  
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;  
QY 1 MNRGVPFRHLILVQLALLPAATQGNKRVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
DB 1 MNRGVPFRHLILVQLALLPAATQGNKRVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYQL 120  
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGITANSSTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
DB 121 LVFGITANSSTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
QY 181 TWTCVILONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLATFVEKLTGSGELMW 240  
DB 181 TWTCVILONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLATFVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360  
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360  
QY 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVCAPPEPKSC 404  
DB 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVCAPPEPKSC 404

RESULT 10  
US-09-243-008-5  
Sequence 5, Application US/09243008  
Publication No. US2004000534A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian et al.  
TITLE OF INVENTION: Redirection of Cellular Immunity by

Receptor Chimeras  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fieh & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243,008  
FILING DATE: 02-Feb-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,176  
FILING DATE: SEPTEMBER 11, 1995  
APPLICATION NUMBER: 08/203,866  
FILING DATE: February 28, 1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: March 6, 1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: March 7, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Karen F. Lech, Ph.D.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/270001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-243-008-5

Query Match 59.5%; Score 2032.5; DB 11; Length 462;  
Best Local Similarity 98.0%; Pred. No. 1.2e-135;  
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;  
QY 1 MNRGVPFRHLILVQLALLPAATQGNKRVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
DB 1 MNRGVPFRHLILVQLALLPAATQGNKRVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYQL 120  
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGITANSSTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
DB 121 LVFGITANSSTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
QY 181 TWTCVILONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLATFVEKLTGSGELMW 240  
DB 181 TWTCVILONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLATFVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360  
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360  
QY 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVCAPPEPKSC 404  
DB 361 LNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVCAPPEPKSC 404



GENERAL INFORMATION:  
APPLICANT: Capon, Daniel J.  
Gregory, Timothy J.  
TITLE OF INVENTION: Adheson Variants  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/097,044A  
FILING DATE: 28-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,918  
FILING DATE: 1-JUN-1995  
APPLICATION NUMBER: 08/236311  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 07/936190  
FILING DATE: 26-AUG-1992  
APPLICATION NUMBER: 07/842777  
FILING DATE: 18-FEB-1992  
APPLICATION NUMBER: 07/250785  
FILING DATE: 28-SEP-1988  
APPLICATION NUMBER: 07/104329  
FILING DATE: 02-OCT-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubic, Jeffrey S.  
REGISTRATION NUMBER: 36,575  
REFERENCE/DOCKET NUMBER: P0444P1C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8228  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 402 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-097-044A-1  
Query Match 59.1%; Score 2017; DB 14; Length 402;  
Best Local Similarity 99.7%; Pred. No. 1.2e-134;  
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60  
1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60  
Qy 1 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
61 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
Db 61 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
Qy 121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
Db 121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
Qy 181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
Qy 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300  
241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300  
Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300

Qy 301 LEATGKLHOBVNLVWRATOLQKNLTCCEVWGPTSPKLMSTKLENKAKVSKKEKPYWV 360  
301 LEATGKLHOBVNLVWRATOLQKNLTCCEVWGPTSPKLMSTKLENKAKVSKKEKPYWV 360  
Db 361 LNPEAGWMOCLLSDSGOVLLESNIKVLPTWSTP 393  
361 LNPEAGWMOCLLSDSGOVLLESNIKVLPTWSTP 393  
RESULT 14  
US-10-103-597A-39  
Sequence 39, Application US/10103597A  
Publication No. US20030096432A1  
GENERAL INFORMATION:  
APPLICANT: Jakobson, Bent Karsten  
TITLE OF INVENTION: Screening Methods  
FILE REFERENCE: 102286.142  
CURRENT APPLICATION NUMBER: US/10/103,597A  
CURRENT FILING DATE: 2002-10-17  
PRIOR APPLICATION NUMBER: PCT/GB00/03579  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: GB 9922352.1  
PRIOR FILING DATE: 1999-09-21  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 39  
LENGTH: 458  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-103-597A-39  
Query Match 59.0%; Score 2015; DB 14; Length 458;  
Best Local Similarity 99.5%; Pred. No. 2e-134;  
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60  
1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60  
Qy 61 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
61 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
Db 61 ILGNQGSFLLTKGPKSLNDRADSRSLMDQGNFPIIKNLKIETSDTYICEVEDQKEEVOL 120  
Qy 121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
Db 121 LVFGILTANSDTHLLOQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
Qy 181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVERSPPLAFVETKLTSGGELMW 240  
Qy 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300  
241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300  
Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKLLPLHLTLPOALPOYAGSGNLTIA 300  
Qy 361 LNPEAGWMOCLLSDSGOVLLESNIKVLPTWSTP 394  
361 LNPEAGWMOCLLSDSGOVLLESNIKVLPTWSTP 394  
Db 361 LNPEAGWMOCLLSDSGOVLLESNIKVLPTWSTP 394

RESULT 15  
US-10-188-444-39  
Sequence 39, Application US/10188444  
Publication No. US20030104635A1  
GENERAL INFORMATION:  
APPLICANT: Jakobson, Bent Karsten  
TITLE OF INVENTION: Screening Methods  
FILE REFERENCE: 102286.142 (CIP)

```

; CURRENT APPLICATION NUMBER: US/10/188,444
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: PCT/GB00/03579
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: GB 9922352.1
; PRIOR FILING DATE: 1999-09-21
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-444-39

Query Match      59.0%; Score 2015; DB 14; Length 458;
Best Local Similarity 99.5%; Pred. No. 2e-134;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60

QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240
DB 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240

QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300

QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300

QY 301 LEAKTGKLOHEVNLVYMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWY 360
DB 301 LEAKTGKLOHEVNLVYMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWY 360

QY 361 LNPEAGMOCCLSDSGQVLLSENIKVLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGQVLLSENIKVLPTWSTPV 394

RESULT 16
US-10-207-655-170
; Sequence 170, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden-Ledbetter, Martha S.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069, 401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 170
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-207-655-170

Query Match      59.0%; Score 2015; DB 14; Length 458;
Best Local Similarity 99.5%; Pred. No. 2e-134;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60
```

```

QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240
DB 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240

QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300

QY 301 LEAKTGKLOHEVNLVYMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWY 360
DB 301 LEAKTGKLOHEVNLVYMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWY 360

QY 361 LNPEAGMOCCLSDSGQVLLSENIKVLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGQVLLSENIKVLPTWSTPV 394

RESULT 17
US-10-151-274-3
; Sequence 3, Application US/10151274
; Publication No. US20030064071A1
; GENERAL INFORMATION:
; APPLICANT: Liltman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geijlendeck, Theo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY
; TITLE OF INVENTION: INTO
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/10/151,274
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US/09/517,605
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-151-274-3

Query Match      58.8%; Score 2007; DB 12; Length 458;
Best Local Similarity 99.2%; Pred. No. 7.4e-134;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNNOIK 60

QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQL 120

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240
DB 181 TWCTVLOQOKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMW 240

QY 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAEARASSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLTLPALPOYAGSGNLTLLA 300
```

Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOYAGSGNLTIA 300  
 Oy 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYV 360  
 Db 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
 Oy 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394  
 Db 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394

## RESULT 18

US-09-891-119A-2

Sequence 2, Application US/09891119A

Publication No. US20040013683A1

GENERAL INFORMATION:

APPLICANT: Maddon, Paul J.

TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4

FILE REFERENCE: 24577-CY-B

CURRENT APPLICATION NUMBER: US/09/891,119A

CURRENT FILING DATE: 2001-06-25

NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn version 3.1

SEQ ID NO: 2

LENGTH: 397

TYPE: PRT

ORGANISM: Human

US-09-891-119A-2

## Query Match

58.6%; Score 2001; DB 11; Length 397;

Best Local Similarity 98.7%; Pred. No. 1.6e-133;

Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60  
 Db 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60  
 Oy 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120  
 Db 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120  
 Oy 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Db 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Oy 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Db 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Oy 181 TWCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVFSFPLAFTVEKLTGSGELMW 240  
 Db 181 TWCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVFSFPLAFTVEKLTGSGELMW 240  
 Oy 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOYAGSGNLTIA 300  
 Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOYAGSGNLTIA 300  
 Oy 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYV 360  
 Db 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
 Oy 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394  
 Db 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394

## RESULT 19

US-08-681-219-27

Sequence 27, Application US/08681219

Publication No. US20020058607A1

GENERAL INFORMATION:

APPLICANT: Takeaki Sato and Junn Yanagisawa

TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN

TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/681,219  
 FILING DATE: 22-Jul-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 0575/48962/JPM/JKM  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 458 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-681-219-27

## Query Match

58.5%; Score 1997; DB 8; Length 458;

Best Local Similarity 98.7%; Pred. No. 3.8e-133;

Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60  
 Db 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60  
 Oy 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120  
 Db 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120  
 Oy 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Db 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Oy 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Db 121 LVFGLTANSPTHLIQGSLTLTLESPPGSSPVQCRSPRKNIOGKTLISVSOLELDSG 180  
 Oy 181 TWCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVFSFPLAFTVEKLTGSGELMW 240  
 Db 181 TWCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVFSFPLAFTVEKLTGSGELMW 240  
 Oy 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOYAGSGNLTIA 300  
 Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOYAGSGNLTIA 300  
 Oy 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYV 360  
 Db 301 LEATGKLHGEVNLVVMKATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
 Oy 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394  
 Db 361 LNPEAGMWQCLLSDSGQVLLSNNIKVLPWTSTPV 394

## RESULT 20

US-09-230-111C-25

Sequence 25, Application US/09230111C

Publication No. US20030203414A1

GENERAL INFORMATION:

APPLICANT: Sato, Taka-Aki

TITLE OF INVENTION: COMPOUNDS THAT INHIBIT INTERACTION BETWEEN

TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF (PDZ/DHR)

CORRESPONDENCE ADDRESS:

```

; TITLE OF INVENTION: DOMAIN AND USES THEREOF
;
; FILE REFERENCE: 48962-A-PCT-US
;
; CURRENT APPLICATION NUMBER: US/09/230,111C
;
; CURRENT FILING DATE: 1999-05-17
;
; NUMBER OF SEQ ID NOS: 33
;
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 25
;
; LENGTH: 458
;
; TYPE: PRT
;
; ORGANISM: human
;
; US-09-230-111C-25

```

Query Match	58.5%	Score 1997:	DB 11;	Length 458;
Best Local Similarity	98.7%;	Pred. No. 3.8e-133;		
Matches 389;	Conservative 1;	Mismatches 4;	Indels 0;	Gaps 0;

[illegible]

QY	61	ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYQL	120
Db	61	ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEYQL	120

```
QY      121 LVFGTANSDTHLLQGSLTTLTLESPGSSPSVQCRRPRKNIQGGKTLVSQLELDQSG 180
      |||||
      121 LVFGITANSDTHLLQGSLTTLTLESPGSSPSVQCRRPRKNIQGGKTLVSQLELDQSG 180
Db
```

[illegible][illegible]

```
QY 301 LEAKTGLHGEVNLVVMRATQIQKNLTCCEWGPSTPKLMSLKLLENKAQKSKREKPYVW 360
DB 301 LEAKTGLHGEVNLVVMRATQIQKNLTCCEWGPSTPKLMSLKLLENKAQKSKREKPYVW 360
```

<b>Qy</b>	<b>361</b>	LNP <sup>EAGM</sup> QC <sup>LLSDSGV</sup> LESN <sup>IKVLPTW</sup> SPV	<b>394</b>
<b>Db</b>	<b>361</b>	LNP <sup>EAGM</sup> QC <sup>LLSDSGV</sup> LESN <sup>IKVLPTW</sup> SPV	<b>394</b>

RESULT 21  
US-10-092-138-25  
; Sequence 25, Application US/10092138  
; Publication No. US20030170723A1

```

; GENERAL INFORMATION:
; APPLICANT: Sato, Taka-Aki
; TITLE OF INVENTION: METHOD OF PREPARING A PROTEIN ARRAY BASED ON
; TITLE OF INVENTION: BIOCHEMICAL PROTEIN-PROTEIN INTERACTION

```

```

; FILE REFERENCE: 030623/CEW/EA
; CURRENT APPLICATION NUMBER: US/1
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 33

```

```

; SEQ ID NO 25
; LENGTH: 458
; TYPE: PRT
;

```

Query Match	58.5%;	Score 1997;	DB 14;	Length 458;
S-10-092-138-25				

1 MNRGVPFRHLLVLQALLPAATGKGVLLGKKGDVLELTCTASQKSIQFHKKNSNQIK 60  
 Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0

Db 1 MNRGVPFRHLLLVLTQLALLPAATQGGKAVLGGKGDTELTCTASQKSSIQFHMNSNQIK 60

Qy 61 ILGNQSFLLTGPSPKLNDRAISRSLMDQGNPPLIIKNLKIENSDTYICEVEDOKEEYQL 120

```
Db      |||||
        |||||
61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPLIIKNLKIEDSDITICEVEDQKEEVQL 120
```

**OY**      121 LVFGLTANSDTHLLQGSLLTTLESPRGSSPPVQCSPRGKNIQGSKTLASVQLLELDGSG 180  
		:		
**DB**      121 LVFGLTANSDTHLLQGSLLTTLESPRGSSPPVQCSPRGKNIQGSKTLASVQLLELDGSG 180

Qy	Db
181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKGEQVEFSFPLAFTVEKLTGSGELMW 2400	181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKGEQVEFSFPLAFTVEKLTGSGELMW 2400

Oy 241 QAEBASSSSWITFDLKNKEVSVKRVTDPLQMGKKLPHLTLPOLPYAGSGNLTIA 3000  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 241 QAEBASSSSWITFDLKNKEVSVKRVTDPLQMGKKLPHLTLPOLPYAGSGNLTIA 3000

```
QY      301 LEAATGKTHQENVLVMMRATQLOKNTLCEVWGPTSPKLTSLKLENKEAKYSKREKPYMV 3600
      |||||
DB      301 LEAATGKTHQENVLVMMRATQLOKNTLCEVWGPTSPKLTSLKLENKEAKYSKREKPYMV 3600
```

Oy	361 LNPEAGMOCLLSDSGQVLESNIKVLPTWSTPV	394
Db	361 LNPEAGMOCCLLSDSGQVLESNIKVLPTWSTPV	394

RESULT 22  
US-10-097-044A-4  
; Sequence 4, Application US/10097044A  
; Publication No. US20030143220A1

/ GENE/PLANT ORIGIN: /  
 / APPLICANT: Capon, Daniel J. /  
 / Gregory, Timothy J. /  
 / TITLE OF INVENTION: Adhesion Variants /

NORTON OF BRUNO: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd

STATE: California  
COUNTRY: USA  
ZIP: 94080

MEDIUM TYPE: 5.25 inch, 360 Kb flo  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/097, 044A  
FILING DATE: 28-May-2002

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,918  
FILING DATE: 1-JUN-1995

FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 07/936190  
FILING DATE: 26-AUG-1992

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8228  
TELEFAX: 415/952-9881



TYPE: amino acid  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 4  
US-10-097-044A-4

Query Match 55.8%; Score 1904; DB 14; Length 434;

Best Local Similarity 99.7%; Pred. No. 1.4e-126;

Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVLGKGDVLTCTASQKSIQFHMKNNSQIKILNQGSLTGTGPKLNDRAISR 83  
DB 56 QGNKVLGKGDVLTCTASQKSIQFHMKNNSQIKILNQGSLTGTGPKLNDRAISR 115  
QY 84 RSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQLVFGLTANSDPHLLQGSLTTL 143  
DB 116 RSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQLVFGLTANSDPHLLQGSLTTL 175  
QY 144 ESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLOQNKKEVFKIDIVVLA 203  
DB 176 ESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLOQNKKEVFKIDIVVLA 235  
QY 204 FOKASSIYKKEGEQVEFSFPLAFVETKLTSGGELMWQABRASSKSWITFDLKNKEYSV 263  
DB 236 FOKASSIYKKEGEQVEFSFPLAFVETKLTSGGELMWQABRASSKSWITFDLKNKEYSV 295  
QY 264 KRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKJHOEVLVVMRATQLO 323  
DB 296 KRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKJHOEVLVVMRATQLO 355  
QY 324 KNLTCFVWGPTSPKMLSLKLENKEAKYSKREKPVWVLANPEAGMWQCLSDSGVLLSEN 383  
DB 356 KNLTCFVWGPTSPKMLSLKLENKEAKYSKREKPVWVLANPEAGMWQCLSDSGVLLSEN 415  
QY 384 IKVLPTWSTP 393  
DB 416 IKVLPTWSTP 425

## RESULT 23

US-09-759-841-6  
Sequence 6, Application US/09759841  
Patent No. US20010039026A1

GENERAL INFORMATION:

APPLICANT: Rickett, Graham A

APPLICANT: Dobbs, Susan

APPLICANT: Petrof, Manoussos

TITLE OF INVENTION: Assay Method

FILE REFERENCE: PCI0348APME

CURRENT APPLICATION NUMBER: US/09/759,841

PRIOR FILING DATE: 2001-01-12

PRIOR APPLICATION NUMBER: GB 0000661.9

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: GB 0000663.5

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: GB 0000659.3

PRIOR FILING DATE: 2000-01-12

NUMBER OF SEQ ID NOS: 6

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 6

LENGTH: 370

TYPE: PRT

ORGANISM: Homo sapiens

US-09-759-841-6

Query Match 55.4%; Score 1891; DB 9; Length 370;

Best Local Similarity 99.7%; Pred. No. 9.3e-126;

Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 KVLGKGDVLTCTASQKSIQFHMKNNSQIKILNQGSLTGTGPKLNDRAISRSL 86  
DB 2 KVLGKGDVLTCTASQKSIQFHMKNNSQIKILNQGSLTGTGPKLNDRAISRSL 61  
QY 87 WDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQLVFGLTANSDPHLLQGSLTTLTLESP 146

DB 62 WDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQLVFGLTANSDPHLLQGSLTTLTLESP 121  
QY 147 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLOQNKKEVFKIDIVVLAFOK 206  
DB 122 PGSSPSVQCRSPRKNIOGKTLVSQLELDQSGTWCTVLOQNKKEVFKIDIVVLAFOK 181  
QY 207 ASSIYKKEGEQVEFSFPLAFVETKLTSGGELMWQABRASSKSWITFDLKNKEYSYRV 266  
DB 182 ASSIYKKEGEQVEFSFPLAFVETKLTSGGELMWQABRASSKSWITFDLKNKEYSYRV 241  
QY 267 TQDPKLOMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKJHOEVLVVMRATQLOKNT 326  
DB 242 TQDPKLOMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKJHOEVLVVMRATQLOKNT 301  
QY 327 TCEVWGPTSPKMLSLKLENKEAKYSKREKPVWVLANPEAGMWQCLSDSGVLLSENKV 386  
DB 302 TCEVWGPTSPKMLSLKLENKEAKYSKREKPVWVLANPEAGMWQCLSDSGVLLSENKV 361  
QY 387 LPTWSTPV 394  
DB 362 LPTWSTPV 369

## RESULT 24

US-10-024-329-32  
Sequence 32, Application US/10024329  
Publication No. US20030157063A1

GENERAL INFORMATION:

APPLICANT: SANHADI, Kamel

APPLICANT: TOURAINE, Jean-Louis

APPLICANT: LEROY, Pierre

APPLICANT: MEHTALI, Majid

TITLE OF INVENTION: Gene therapy using anti-gp41 antibody and cd4 immunoadhesin

FILE REFERENCE: 109993

CURRENT APPLICATION NUMBER: US/10/024,329

PRIOR FILING DATE: 2001-12-21

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 32

LENGTH: 448

TYPE: PRT

ORGANISM: human scd4

US-10-024-329-32

Query Match 55.0%; Score 1877.5; DB 14; Length 448;

Best Local Similarity 97.0%; Pred. No. 1.1e-124;

Matches 382; Conservative 0; Mismatches 3; Indels 9; Gaps 8;

QY 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGDVLTCTASQKSIQFHMKNNSQIK 60  
DB 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGDVLTCTASQKSIQFHMKNNSQIK 59  
QY 61 ILNQGSLTGTGPKLNDRAISRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQL 120  
DB 60 ILNQGSLTGTGPKLNDRAISRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVQL 117  
QY 121 LVFGLTANSDPHLLQGSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
DB 118 LVFGLTANSDPHLLQGSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 174  
QY 181 TWCTVLOQNKKEVFKIDIVVLAFOKASSIYKKEGEQVEFSFPLAFVETKLTGS-ELMW 240  
DB 175 TWCTVLOQNKKEVFKIDIVVLAFOKASSIYKKEGEQVEFSFPLAFVETKLTGS-ELMW 231  
QY 241 QABRASSKSWITFDLKNKEYSYRVATQDPKLOMGKKLPHLTLPOALPOYAGSGNLTLA 300  
DB 232 QABRASSKSWITFDLKNKEYSYRVATQDPKLOMGKKLPHLTLPOALPOYAGSGNLTLA 291  
QY 301 LEAKTGKJHOEVLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPVV 360  
DB 292 LEAKTGKJHOEVLVVMRATQLOKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPVV 351

QY 361 LNPEAGMOCCLSDSGVILLESNIKVLPTWSTPV 394  
DB 352 LNPEAGMOCCLSDSGVILLESNIKVLPTWSTPV 385

## RESULT 25

US-09-939-537-33

Sequence 33, Application US/09939537

Publication No. US20030138410A1

GENERAL INFORMATION:

APPLICANT: Seed, Brian

Banapur, Babak

Romeo, Charles

Kolanus, Waldemar

TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INJECTED

CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS

NUMBER OF SEQUENCES: 53

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark &amp; Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/939,537

FILING DATE: 24-Aug-2001

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/284,391

FILING DATE: 02-AUG-1994

APPLICATION NUMBER: 08/195,395

FILING DATE: 14-FEB-1994

APPLICATION NUMBER: 07/847,566

FILING DATE: 06-MAR-1992

APPLICATION NUMBER: 07/665,961

FILING DATE: 07-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Elbing, Karen L.

REGISTRATION NUMBER: 35,238

REFERENCE/DOCKET NUMBER: 00786/247001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX: &lt;Unknown&gt;

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 254 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 33:

US-09-939-537-33

Query Match

Best Local Similarity 98.0%; Pred. No. 8.5e-87; Length 254;

Matches 249; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 400 EPKSCDKTHTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVAVVSHEDPEVKF 454

DB 1 EPKSCDKTHTCPCPAPPELLGSPVFLFPKPKDTLMISRTPEVTCVAVVSHEDPEVKF 60

QY 455 NMVYDGVVNAKTKPRBEQYNSTYRVVSVLTVLIHQMLNGKCYKCVSKALPAPLEKT 514

DB 61 NMVYDGVVNAKTKPRBEQYNSTYRVVSVLTVLIHQMLNGKCYKCVSKALPAPLEKT 120

QY 515 ISKAKGPRREQVYTLPPSRDELTKNQVSLTCLVKGYFPPSDIAVWESNGQPENNYKTP 574

DB 121 ISKAKGPRREQVYTLPPSRDELTKNQVSLTCLVKGYFPPSDIAVWESNGQPENNYKTP 180

QY 575 PVLDSGSPFLYSKLTVDKSRWQGNVFSQVMEALNNHTYOKSLISPGIQLDETCAE 634

DB 181 PVLDSGSPFLYSKLTVDKSRWQGNVFSQVMEALNNHTYOKSLISPGIQLDETCAE 240

QY 635 AODGELDGLWTTDP 648

DB 241 AODGELDGLWTTDP 254

## RESULT 26

US-10-363-427-18

Sequence 18, Application US/10363427

Publication No. US2003019538A1

GENERAL INFORMATION:

APPLICANT: Medexgen Inc.

APPLICANT: CHUNG, Yong Hoon

APPLICANT: HAN, Ji Woong

APPLICANT: LEE, Hye Ja

APPLICANT: CHOI, Eun Yong

APPLICANT: KIM, Jin Mi

APPLICANT: YIM, Soo Bin

TITLE OF INVENTION: Concatameric Immunoadhesion

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/363,427

CURRENT FILING DATE: 2003-02-28

NUMBER OF SEQ ID NOS: 52

SOFTWARE: Kopatentia 1.71

SEQ ID NO 18

LENGTH: 617

TYPE: PRT

ORGANISM: Homo sapiens

US-10-363-427-18

Query Match

Best Local Similarity 49.0%; Pred. No. 1.7e-84;

Matches 317; Conservative 52; Mismatches 137; Indels 141; Gaps 24;

QY 35 DIVELTCTASQKSIQFHWKNSNQIKLGNQGSFLTKGPKLNDRAHSRLMDQGNFPL 94

DB 55 DDKMEKTSDDKKIAQFRKEKE-----TFKEDTYKLFK-----NGTL 92

QY 95 IIKNLKIEDSDTYICEVEQK-EVQVLVFGLTANSDTLLDQO-----SLTFL 142

DB 93 KIKHLKTDQDIIYKVSIVDTGKRVLEKIFDLK-----IQERSKPKISWTCINTTLL 145

QY 143 LSSPSSPSVOCSPBGRKNIQGGKTLVSQLELDQSGTGTCTVLQNKVEFKIDIVL 202

DB 146 CEVNGTDBELNL-----YDGGKHLKLSQVIT-THKNTLSL-----SAFK-CT 187

QY 203 AFQKASIVYKKEGQVEFSPLAFTVEKLTGSGELM----- 239

DB 188 AGNKVS-----KESSVEPVSCP-----KEITVALFTWGLAGDINDIPSFQWDDIDDI 237

QY 240 -WQERASSSSKSWITFD-----LAKKEVSKRYTQPKLQMGKPLPHLTL 284

DB 238 KW--EKTSDDKKIAQFRKEKEKTFKEDTYKLFKNGTLKIKHLKTDQ--ODIYKVSIVDT- 292

QY 285 PQLPQVYASGNLTALAEKGTGLHGVNLVVMRATOLQKNTLCEVWGTSPPLMLSLKL 344

DB 293 -----KKNVLEKIFDL-----KIQERSKPKISWTCINTTLLCEVNGTDBE--LNLVQ 340

QY 345 ENKEAKVSKR-KPVVNLNPEAGMOCCLSDSGVILLESNIKVLPTWSTPVPCPAPPEK 403

DB 341 DGHKLTLGQRYVITHKMTTSLA-KFKC--TAGNKVSKSSV-----PVSCPA-EPKS 389

QY 404 CDKTHTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVAVVSHEDPEVKRWY 458

DB 390 CDKTHTCPCPAPPELLGSPVFLFPKPKDTLMISRTPEVTCVAVVSHEDPEVKRWY 449

QY 459 DGVVNAKTKPRBEQYNSTYRVVSVLTVLIHQMLNGKCYKCVSKALPAPLEKTSK 518

Db 450 DGVEVHNAKTPREEOYNSTYRVVSVLTVCCHQDWLNGKEYKCKVSNKALPAPIEKTISKA 509  
|||  
Qy 519 KGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLD 578  
|||  
Db 510 KGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLD 569  
|||  
Qy 579 SDGSFPLYSKLTVDKSRMGOGNVFSQSVMEALHNHYTQKSLSLSPG 625  
|||  
Db 570 SDGSFPLYSKLTVDKSRMGOGNVFSQSVMEALHNHYTQKSLSLSPG 616  
|||

## RESULT 27

US-10-363-427-22  
; Sequence 22, Application US/10363427  
; Publication No. US20030195338A1  
; GENERAL INFORMATION:  
; APPLICANT: MedexGen Inc.  
; APPLICANT: CHUNG, Yong Hoon  
; APPLICANT: HAN, Ji Woong  
; APPLICANT: LEE, Hye Ja  
; APPLICANT: CHOI, Eun Yong  
; APPLICANT: KIM, Jin Mi  
; APPLICANT: YIM, Soo Bin  
; TITLE OF INVENTION: Concatametric Immunoadhesion  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/363,427  
; CURRENT FILING DATE: 2003-02-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Kopatentin 1.71  
; SEQ ID NO 22  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-363-427-22

Query Match 38.5%; Score 1313.5; DB 14; Length 617;  
Best Local Similarity 49.0%; Pred. No. 1.7e-84;

Matches 317; Conservative 51; Mismatches 138; Indels 141; Gaps 24;

Qy 35 DTVELTCTAQQKSIQFHMKNSNOIKILGNQSFILTKGPKLNDRADRSRLMDQGNFPL 94  
|||  
Db 55 DDIMKETSIDKCKIAQFRKEKE-----TFEKDYTLFK-----NGTL 92  
|||  
Qy 95 IIRKLKIEDSPITYCEVEDQK-EEVQQLVFGLTANSSTHLLQCG-----SLTIT 142  
|||  
Db 93 KIRKLKTDODIYKVSIVDTGKRVLEKIFDLK-----IQERVSKPKISWTCINTTLLT 145  
|||  
Qy 143 LESPSSPSVQGRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNGKVEFKIDIVL 202  
|||  
Db 146 CEVWNGTIDPELNL-----YQDKHLKLSORVI--THKWTLSL-----SAKFK---CT 187  
|||  
Qy 203 AFQKASSIVYKEGEQVEFSPFLAFTVEKLTGSGELM----- 239  
|||  
Db 188 AGNKVS-----KESSEVENVSCP-----KNITNALETWGAQGDINDLIPSRQMSDDIDDI 237  
|||  
Qy 240 -WQAEKRSSSKSMWTFD-----LKKKEVSVKRVTDPPKLGNGKKLPLHLTL 284  
|||  
Db 238 KM--EKTSDDKCKIAQFRKEKETFEKDYTLFKNGTILKIRKLKTD--ODIYKVSIVDT- 292  
|||  
Qy 285 PQALPQVAGSGLTLALAKTGKHOEYNLVVMKATOLQKNLTCEVWNGPTSPKLMLSLKL 344  
|||  
Db 293 -----KGNVLEKIDL--KIQERVSKPKISWTCINTTILCEVWNGTIDPE--LNLVQ 340  
|||  
Qy 345 ENKEAKVSKRE-KPVWVNLPEAGMMQCLSDSGVLLSNIKVLPTWSTVPVCPAPPEPKS 403  
|||  
Db 341 DGKILKLSQRIYTHKMTTSLSA-KFKC--TAGNKVSKSSSVE-----PVSCPA--EPKS 389  
|||  
Qy 404 CDKHTHC-----PELLGSPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNWYV 458  
|||  
Db 390 CDKHTHCPCPCAPPELLGSPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNWYV 449  
|||  
Qy 459 DGVEVHNAKTPREEOYNSTYRVVSVLTVCCHQDWLNGKEYKCKVSNKALPAPIEKTISKA 518  
|||

Db 450 DGVEVHNAKTPREEOYNSTYRVVSVLTVCCHQDWLNGKEYKCKVSNKALPAPIEKTISKA 509  
|||  
Qy 519 KGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLD 578  
|||  
Db 510 KGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLD 569  
|||  
Qy 579 SDGSFPLYSKLTVDKSRMGOGNVFSQSVMEALHNHYTQKSLSLSPG 625  
|||  
Db 570 SDGSFPLYSKLTVDKSRMGOGNVFSQSVMEALHNHYTQKSLSLSPG 616  
|||

## RESULT 28

US-10-363-427-14  
; Sequence 14, Application US/10363427  
; Publication No. US20030195338A1  
; GENERAL INFORMATION:  
; APPLICANT: MedexGen Inc.  
; APPLICANT: CHUNG, Yong Hoon  
; APPLICANT: HAN, Ji Woong  
; APPLICANT: LEE, Hye Ja  
; APPLICANT: CHOI, Eun Yong  
; APPLICANT: KIM, Jin Mi  
; APPLICANT: YIM, Soo Bin  
; TITLE OF INVENTION: Concatametric Immunoadhesion  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/363,427  
; CURRENT FILING DATE: 2003-02-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Kopatentin 1.71  
; SEQ ID NO 14  
; LENGTH: 437  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-363-427-14

Query Match 37.4%; Score 1275.5; DB 14; Length 437;  
Best Local Similarity 77.8%; Pred. No. 5.2e-82;

Matches 253; Conservative 18; Mismatches 35; Indels 19; Gaps 7;

Qy 307 KLEHVEYNLVVMKATOLQKNLTCEVWNGPTSPKLMLSLENKEAKVSKRE-KPVWVNLPEA 365  
|||  
Db 125 KIGERVSKPKISWTCINTTILCEVWNGTIDPE--LNLVQDKHLKLSORVI--THKMTTSLSA 182  
|||  
Qy 366 GMMQCLSDSGVLLSNIKVLPTWSTVPVCPAPPEKSCDKHTC-----PELLGSPSV 420  
|||  
Db 183 -KFKC--TAGNKVSKSSSVE-----PVSCPA--EPKSCDKHTCPCPCAPPELLGSPSV 231  
|||  
Qy 421 LPPKPKDPTLMSRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTPREEOYNSTYR 480  
|||  
Db 232 LPPKPKDPTLMSRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTPREEOYNSTYR 291  
|||  
Qy 481 VSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 540  
|||  
Db 292 VSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 351  
|||  
Qy 541 QVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLDSDGSFPLYSKLTVDKSRMGOGN 600  
|||  
Db 352 QVSLTCLVKGFPSDIAVEMESNGQPENNYKTTPVLDSDGSFPLYSKLTVDKSRMGOGN 411  
|||  
Qy 601 VFSCSVMEALHNHYTQKSLSLSPG 625  
|||  
Db 412 VFSCSVMEALHNHYTQKSLSLSPG 436  
|||

RESULT 29  
US-09-935-868-8  
; Sequence 8, Application US/09935868  
; Patent No. US20020164690A1  
; GENERAL INFORMATION:  
; APPLICANT: Regeneron Pharmaceuticals, Inc  
; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using  
; FILE REFERENCE: REG 203D



```

; TITLE OF INVENTION: AND USING
; FILE REFERENCE: REG 203-B-US
; CURRENT APPLICATION NUMBER: US/10/282,162
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: 09/787,835
; PRIOR FILING DATE: 1999-09-22
; PRIOR APPLICATION NUMBER: PCT/US99/22045
; PRIOR FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-282-162-8

```

```

Query Match      37.3%; Score 1275; DB 14; Length 592;
Best Local Similarity 47.3%; Pred. No. 8,4e-82;
Matches 308; Conservative 53; Mismatches 160; Indels 130; Gaps 19;

```

```

QY 20 PAATGKRVVLGKGDVLELTCTASQ-KKSIQFHKNKSNQIKILGNQSGFLTKGPKSLND 78
DB 26 PAQAVARGLVLTSPDSDVTLTCPCVGEPRDNATVHW-----VLRKPA----- 66
QY 79 RADSRRLMDQGNFPLIIKNLIKIEDSDTYICE-----VEDQKEEVLVFGLT 126
DB 67 -AGSHSRWAGKGRLLRSVQYLDHSGNYSCYRAGRAGTVHLLVDVPEEPQLSCFRKS 125
QY 127 ANSDTHLLQGSGLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSGTWCTV 186
DB 126 PLSN-----VVCSEGRSTPSLTTKA-----VLLVRKQNSPAEDFQFPC 165
QY 187 LONQKVEFKIDIVLAFQKASIIYKKKEGVEFSPFLAFVETKLTSGEL----- 238
DB 166 QYSGSQKFSQCLAVPEGDSFYIIVSMCVASSVSGSKETQTFQ--GCGILQDPRPANI 222
QY 239 -----W-----MQAERASSSKSM-ITFDLKNKEVSVGRVTPDPLQMGKTLPLTLT 283
DB 223 TVTAVARNRYMLSVTWODRPHSNSSFYRLRFELRYBARSKFT---TWVVKDLQNHCV 278
QY 284 LPOLPOLYAGSGNLTLLAEAKTGKLDQEVNLVVMRATOLQKLTCEWGWPTSPKLMISLK 343
DB 279 IH-----DAWSGLRH-----VVOJRA--QEFQGGSEWSEWSPRAMGTPW 315
QY 344 LENK-----EAKVSKREKPVWVNLNPEAGMOCILSDSGVLLSNIKULPTMTSTVPCAP 399
DB 316 TESRRPRAENEVS--TPMQALTTNKDDNILFRDS-----ANATSLPVQD-----AG 360
QY 400 EPKSCDKTHTC-----PELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
DB 361 EPKSCDKHTTCPCAPARLLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 420
QY 455 NMVYDGVENNAKTKPREEOYNSTYRVVSVLTVLHQDLNKGKEYKCKVSKALPAPIEKT 514
DB 421 NMVYDGVENVNAKTKPREEOYNSTYRVVSVLTVLHQDLNKGKEYKCKVSKALPAPIEKT 480
QY 515 ISKAKGOREPOVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTTP 574
DB 481 ISKAKGOREPOVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTTP 540
QY 575 PVLDSGSGFLYSKLTVDKSRMOQGNVFSQVMEHALNHNHTOKSLSLSPG 625
DB 541 PVLDSGSGFLYSKLTVDKSRMOQGNVFSQVMEHALNHNHTOKSLSLSPG 591

```

```

RESULT 32
US-09-815-108-22
; Sequence 22, Application US/09815108
; Patent No. US20020009776A1
; GENERAL INFORMATION:
; APPLICANT: Sarris, Christiaan M.
; APPLICANT: Sharon, Wu X.
; APPLICANT: Xia, Min

```

```

; APPLICANT: Boone, Thomas Charles
; APPLICANT: Covey, Todd
; TITLE OF INVENTION: Fibroblast Growth Factor Receptor-Like Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 99-513-A
; CURRENT APPLICATION NUMBER: US/09/815,108
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/191,379
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: murine FGFR-L
; OTHER INFORMATION: extracellular domain-Fc fusion polypeptide
US-09-815-108-22

```

```

Query Match      37.1%; Score 1265; DB 9; Length 594;
Best Local Similarity 44.2%; Pred. No. 4.3e-81;
Matches 311; Conservative 55; Mismatches 132; Indels 206; Gaps 22;

```

```

QY 11 LVLVQLALPLPA-----TQGNKV-----LQK-----KGDVLELTCTASQKSI 49
DB 7 LLLILLGALISAEARPPRMADKVYRQVARIQRTVRLQCPYEGDPEPLTMTKQDRTI 66
QY 50 QFHKNKSNQIKILGNQSGFLTKGPKSLNDRADSRRLMDQGNFPLIIKNLIKIEDSDTYIC 109
DB 67 HSGW-----SRFVLPQG-----LKVEAEADAGVYVC 95
QY 110 EVDQKEEVLVPGLTANS DTHLQGSGLTTLTLESP-PESS----- 150
DB 96 KATNG-----FGSLSVYVTLIM-DDISPKESPPGSGSGGQEDPASQOWARPT 146
QY 151 -PSVOCR-----SPRGKNIQGGK-----TLSVSQ 173
DB 147 QPSMRARRVILARPGSSVRLKCVASGHRPRDMMKDDOTLTHLSEHKKKXTLSLN 206
QY 174 LEIQDSGTWTCTYLQNKQVE--FKIDIVLAFQKASIIYKKKEGVEFSPFLAFVETK 231
DB 207 LKPEDSGKYTCRVSNKAGAINATYKVDVIGRTSKSPVLTGTHPVNTTVDEGCTTSFQCK- 265
QY 232 LTGSGELMWAERASSSKSITFDLKNKEVSVGRVTPDPLQMGKTLPLTLPOLPOLY 291
DB 266 -----VRSDVK-----PVIQWLKRV-----EY 282
QY 292 AGSGNLTLLAEAKTGKLDQEVNLVVMRATOLQKLTCEWV-----GPTSPKLMISLKLENK 347
DB 283 GSEGRHNSTIDVGQK-----FVVLV-----TGVMSRPDGSYLNKLLISRAQD- 327
QY 348 EAKVSKREKPVWVNLNPEAGMOCILSDS--GQVLLSNIKULPTMTSTVPCAPPEKSCDK 406
DB 328 -----DAGNYICIGANTMGYSFRSAFLTVLPDPKPGPRMA--SSSVDK 369
QY 407 THTC-----PELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKNMVYDGV 461
DB 370 THTCPCPAPARLLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKNMVYDGV 429
QY 462 EVNNAKTKPREEOYNSTYRVVSVLTVLHQDLNKGKEYKCKVSKALPAPIEKTISAKQ 521
DB 430 EVNNAKTKPREEOYNSTYRVVSVLTVLHQDLNKGKEYKCKVSKALPAPIEKTISAKQ 489
QY 522 PREPOVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTTPVLDSDG 581
DB 490 PREPOVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTTPVLDSDG 549
QY 582 SFFLYSKLTVDKSRMOQGNVFSQVMEHALNHNHTOKSLSLSPG 625
DB 550 SFFLYSKLTVDKSRMOQGNVFSQVMEHALNHNHTOKSLSLSPG 593

```

RESULT 33  
US-10-229-584-22  
; Sequence 22, Application US/10229584  
; Publication No. US2003087384A1  
; GENERAL INFORMATION:  
; APPLICANT: Sarris, Christiaan M.  
; APPLICANT: Sharon, Mu X.  
; APPLICANT: Xia, Min  
; TITLE OF INVENTION: Fibroblast Growth Factor Receptor-Like Molecules and  
; TITLE OF INVENTION: Uses Thereof  
; FILE REFERENCE: 99-513-F  
; CURRENT APPLICATION NUMBER: US/10/229,584  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 09/815,108  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/191,379  
; PRIOR FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: murine FGFR-L  
; OTHER INFORMATION: extracellular domain-Fc fusion polypeptide  
US-10-229-584-22

Query Match 37.1%; Score 1265; DB 14; Length 594;  
Best Local Similarity 44.2%; Pred. No. 4.3e-81;  
Matches 311; Conservative 55; Mismatches 132; Indels 206; Gaps 22;

QY 11 LVLVLALALPAA-----TQGNKV-----LGR-----KQDVELTCTASQKSI 49  
DB 7 LLLLLLALPSAEARPPPMADKVPVQYARLGRVRLQCPVGEPPPLTWTMGKRTI 66  
QY 50 QPFWKNSNQIKILGNQSGFLTKGPSKLNDRASRSRLMDQGNFPLIKNLKIEDSPYIC 109  
DB 67 HSGH-----SRRFLPQG-----LKKVEADAGVVC 95  
QY 110 EVEDQKEEVLVFGLTANSDFLLQGSITLTLESF-PGSS----- 150  
DB 96 KATNG-----FGSLSVNYTLIM-DisPKESPQSGSGQGPASQOMABRFT 146  
QY 151 -PSVQCR-----SPRKNIOGK-----TLVSQ 173  
DB 147 QPSTMRRTVJARPVGSSVRLKCVASGHPRPDIIMMKDQTLTLLEASERKKWTLSLN 206  
QY 174 LEIADSGTWTCTVLQNOQKVE--FKIDIVLAFQKASSIVYKKEGEVSEFPPLAFTVEK 231  
DB 207 LKPEDSGKTCYCNKAGALNATYKVDVIQRTSKPVLGTHTVNTTVDGCTTSFQCK- 265  
QY 232 LTGSGELMWAERASSKSWITPDLKNEKVSVKRVTDPRKLGKGLPLHLTLPLQALPOY 291  
DB 266 -----VRSIVK-----PVLQMLKRV-----EX 282  
QY 292 AGSGLTLALEAKTKGLHGEVNLVYMARATOLQKLTGEVW-----GPSPTMLSLKLENK 347  
DB 283 GSEGRHNSITDVGQK-----FVLLP-----TDGWSRPDGSYLNKLLSLRARD- 327  
QY 348 EAKVSKREKPVWVLPNPAAGMOCCLSDS-GQVLLSNIKYLPTWSTVPAPPEPKSCDK 406  
DB 328 -----DAGMYICIGANTMGYSRPSALTYLPPDKPGRPMA-----SSVDK 369  
QY 407 THTC-----PELLGSPVFLPPPKDPTLMSRTPEVTQVVDVSHDEPEVKFNWYDGV 461  
DB 370 THTCPAPBELLGSPVFLPPPKDPTLMSRTPEVTQVVDVSHDEPEVKFNWYDGV 429  
QY 462 EVNNAKTRPEEQVNSYRVVSVLTVLHOMLNKEKCKVSNKALPAPLEKTSKAKGQ 521  
DB 430 EVNNAKTRPEEQVNSYRVVSVLTVLHOMLNKEKCKVSNKALPAPLEKTSKAKGQ 489  
QY 522 PREQVYTLPSRDELTKNQVSLTCLVKGYFSPDIAVEMESNGQPENNYKTTTPVLSDG 581

DB 490 PREQVYTLPSRDELTKNQVSLTCLVKGYFSPDIAVEMESNGQPENNYKTTTPVLSDG 549  
QY 582 SFPLYSKLTVDKSRMQQGNVFCSSVMHEALNHYTQKSLISLSPG 625  
DB 550 SFPLYSKLTVDKSRMQQGNVFCSSVMHEALNHYTQKSLISLSPG 593

RESULT 34  
US-09-910-600-16  
; Sequence 16, Application US/09910600  
; Publication No. US20030036631A1  
; GENERAL INFORMATION:  
; APPLICANT: Longphre, Malinda  
; APPLICANT: Chang, Han  
; APPLICANT: Whitney, Gena  
; TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF  
; FILE REFERENCE: D0003NP  
; CURRENT APPLICATION NUMBER: US/09/910,600  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/220,139  
; PRIOR FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 779  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: L3-hig  
US-09-910-600-16

Query Match 37.1%; Score 1265; DB 10; Length 779;

Best Local Similarity 45.0%; Pred. No. 6.2e-81;  
Matches 304; Conservative 43; Mismatches 136; Indels 192; Gaps 19;

QY 8 RHLLVLALALPAA--ATQGNKVVL-GKKQDVELTCTASQKSIQFHWKNSNQIKILG 63  
DB 239 RDLVISISDNTPALPEPQGNVPIYLEAKQGPLLLCAADQPPATLSW----- 288  
QY 64 NQGSFLTKGPSKLNDRASRSRLMDQGNFPLIKNLKIEDSPYICEVEDQ----- 114  
DB 289 -----VLQNRVLSSSHMPGPRPLGLPLPGKADSGRYTCRAENRSGQORALD 337  
QY 115 -----KEEVOLVLF-----GLTANSDFLLQGSITLTLESFPGSSPSVQCS 157  
DB 338 LSVQYPPENLRVWVQANFTVLENLNGTSLPVLGQSCLVCSVTHSSPPA----- 388  
QY 158 PRGKNIOGKTSVSO-----LEIADSGTWTCTVLQNOQKVEFKIDIVLVLAFOK 206  
DB 389 RLSTWQKQVLSFSPQSPDPGVLELPRVQVHEGEFTCHAR----- 428  
QY 207 ASSIVYKKEGEVSEFPPLAFTVEKLTGSGELW-----WQERASSKSWITPDL 256  
DB 429 -----HPLGSHVYSLSLVHYS-PKLTPGSCSWEAEGLCSCSSQASPAISRLMWL- 478  
QY 257 KKEKVSVKRVTDPRKLGKGLPLHLTLPLQALPOYAGS-GNITLMAEATKGLHGEVNLV 315  
DB 429 -----HPLGSHVYSLSLVHYS-PKLTPGSCSWEAEGLCSCSSQASPAISRLMWL- 478  
QY 479 -GELLLEGNSQDSF-----EVTPESSAGPWANSSLSLH-----GGISSGL- 517  
DB 316 VNRATOLQKNLTCCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAAGMOCCLSDS 375  
QY 518 -----RLRCEANVNGAOSGSLQLPDKG-----LISD- 546  
DB 376 GQVLLSNIKYLPTWSTVPAPPEPKSCDKTHTC-----PELLGSPVFLPPPKDPTL 430  
QY 547 -----PEPKSCDKTHTCPPCAPPEBEGAPSVFLPPPKDPTL 583  
QY 431 MISTRTPEVTQVVDVSHDEPEVKFNWYDGVVNAKTRPREQVNSYRVVSVLTVLH 430  
DB 584 MISTRTPEVTQVVDVSHDEPEVKFNWYDGVVNAKTRPREQVNSYRVVSVLTVLH 643  
QY 491 DWLNGREYKCKVSNKALPAPLEKTSKAKGQPREQVYTLPSRDELTKNQVSLTCLVKG 550

|||||  
Db 644 DMLNGKEYKCKVSKNKPALPALEKTIKSKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKG 703  
Qy 551 FYPSDIAVEMESNQGPNNNYKTTTPVLDSDGSFLYYSKLTVDKSRMOQGNVFCSCVWHEA 610  
Db 704 FYPSDIAVEMESNQGPNNNYKTTTPVLDSDGSFLYYSKLTVDKSRMOQGNVFCSCVWHEA 763  
Qy 611 LHHNYTKSLSLSPG 625  
Db 764 LHHNYTKSLSLSPG 778

## RESULT 35

US-09-910-600-30  
; Sequence 30, Application US/09910600  
; Publication No. US20030036631A1  
; GENERAL INFORMATION:  
; APPLICANT: Longphre, Malinda  
; APPLICANT: Chang, Han  
; APPLICANT: Whitney, Gena  
; TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF  
; FILE REFERENCE: D0003ND  
; CURRENT APPLICATION NUMBER: US/09/910,600  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/220,139  
; PRIOR FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 779  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: L3-hig  
US-09-910-600-30

Query Match 37.1%; Score 1265; DB 10; Length 779;  
Best Local Similarity 45.0%; Pred. No. 6.2e-81;  
Matches 304; Conservative 43; Mismatches 136; Indels 192; Gaps 19;  
Qy 8 RHLLVLQALLPA--ATQGNKVLL-GKKGDVLELTCTASOKSIQFHWKNSNQIKLG 63  
Db 239 RDLVISISRDNTPALPEPQGNVPLBAQKQFLRLCAADSPATLSW----- 288  
Qy 64 NQGSFLTGPSKLDNRADSRSLMDQGNFLLIKLKIETSDTYICEVEDQ----- 114  
Db 289 -----VLONRVLSSSHHPGPRPLGELPRVKAGDSGRYTCRAENLIGSQORALD 337  
Qy 115 -----KEEVOLVPE-----GLTANSDTHLLOGOSLTL-----TLESPPGSSPSVQCRS 157  
Db 338 LSVQYPPENLKVMSQANRYLENLNGTSLPVEGOSLCLVCTHSSPPA----- 388  
Qy 158 PRGNKIQQGKTLVSQ-----LELDQSGTWTCTVLONOKKVEFKIDIVLVAQK 206  
Db 389 -RLSWGQGVLSRPSQSDPGVLELPRVQVEHGEFTCHAR----- 428  
Qy 207 ASSIYVKKEGQVESFPLATVEKLTGSGELM-----MOAERASSKSWITFDL 256  
Db 429 -----HPLGSOHVLSLSVHYS-PLLLGPSCSWEABGLHSCSSQASAPSLRWML----- 478  
Qy 257 KKEKESVVRVTPQDPLQWKKLPLHLTLPOALPOYAGS-GNLTALAEKTKLQOEVLV 315  
Db 479 -GEBLEBNSQDSF-----EVTSPSAGWMANSSLSLH---GGLSSGL--- 517  
Qy 316 VWRATQLOKNTCEVWGPTSPKMLSLKLENKEAKVSKREKRVWVLANPEAGMOCLLSDS 375  
Db 518 -----RLRCEAMVNHGAQSGSIQLDPKKG-----LISD- 546  
Qy 376 GOVLLESNIKVLPTMTSTVPCRPAPKSCDKTHTC-----PELLGSPVFLFPPKPKDTL 430  
Db 547 -----PEPKSCDKTHTCPPCPAPFPGAPSVFLFPPKPKDTL 583  
Qy 431 MISRPEVTCVVVDVSHEDPEVKFNNVYDVGVEVHNAKTKPREEQNGNSTRYRVVSVLTVLIHQ 490

|||||  
Db 584 MISRPEVTCVVVDVSHEDPEVKFNNVYDVGVEVHNAKTKPREEQNGNSTRYRVVSVLTVLIHQ 643  
Qy 491 DMLNGKEYKCKVSKNKPALPALEKTIKSKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKG 550  
Db 644 DMLNGKEYKCKVSKNKPALPALEKTIKSKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKG 703  
Qy 551 FYPSDIAVEMESNQGPNNNYKTTTPVLDSDGSFLYYSKLTVDKSRMOQGNVFCSCVWHEA 610  
Db 704 FYPSDIAVEMESNQGPNNNYKTTTPVLDSDGSFLYYSKLTVDKSRMOQGNVFCSCVWHEA 763  
Qy 611 LHHNYTKSLSLSPG 625  
Db 764 LHHNYTKSLSLSPG 778

## RESULT 36

US-10-207-655-348  
; Sequence 348, Application US/10207655  
; Publication No. US20030118592A1  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Hayden-Ledbetter, Martha S.  
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
; FILE REFERENCE: 390069,401C1  
; CURRENT APPLICATION NUMBER: US/10/207,655  
; CURRENT FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 426  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 348  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion polypeptide  
US-10-207-655-348

Query Match 37.0%; Score 1262; DB 14; Length 504;  
Best Local Similarity 46.7%; Pred. No. 5.7e-81;  
Matches 302; Conservative 42; Mismatches 120; Indels 182; Gaps 20;  
Qy 1 MNRGVPPRHLLVLQALLPAATQGNKVLLGKKGDVLELTCTASOKSIQFHW---KNSN 57  
Db 19 MSRGVDIQ-----MTQTSSLSASIGDRITICRASQDIRNLNMYQKPDG 65  
Qy 58 QIKLGNQGSFLLTKGPSKLDNRADSRSLMDQGNFLLIKLKIETSDTYICEVEDQKE 116  
Db 66 TVKXL-----IYYT-----SRLHSGVPSRFGSGSGSDYSLTIANLQPEDIATYFCQ----- 112  
Qy 117 EVOLLVNGLTANSDTHLLOGOSLTLTLESPPGSSPSVQCRSPRGKNIQQGKT---LSVSQ 173  
Db 113 -----QNTLPLWTF-----GGGKVLTKRELQGGSGGGGGGGSIDE 151  
Qy 174 LELODSGTWTCTVLONOKKVEFKIDIVLVAFOKASIVYKKEGQVESFPLATVEKLT 233  
Db 152 VOLQOQSP-----ELV-----KQASMSCAASG---YSF-TGIVN--- 183  
Qy 234 GSGELMWQAEPRASSKSWITFDLKNKEVSVKRVYTPDKLOMGKKLPLHLTLPOALPOYAG 293  
Db 184 -----WLKQSHGKNLEWIGLINPYKGLT-----TYNQKFK----- 213  
Qy 294 SGNLTLLAEKTKLQOEVLVWVRATQLOKNTLCEVWGPTSPKMLSLKLENKEAKVSK 353  
Db 214 -GRATLTVDKSSSTAYME-----LSTLSSEDSAVYYCA 245  
Qy 354 REK-----PYVNLNPEAGMOCLLSDSGVLESNIKVLPTWSTPVPAPPEKSC 404  
Db 246 RSGYIGSDMYFDVW---GAGTTVYSSIQ-----PEKSC 277  
Qy 405 DKTHTC-----PELLGSPVFLFPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFNNVYD 459  
Db 278 DKTHTCPPCPAPFELGGSPSVFLFPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFNNVYD 337

```

QY 460 GVEVNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAK 519
DB 338 GVEVNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAK 397
QY 520 GQREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDG 579
DB 398 GQREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDG 457
QY 580 DGSFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 625
DB 458 DGSFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 503

RESULT 37
US-10-207-655-345
; Sequence 345, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207.655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 345
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion polypeptide
US-10-207-655-345

Query Match          36.9%; Score 1260; DB 14; Length 543;
Best Local Similarity 46.4%; Pred. No. 8.7e-81;
Matches 299; Conservative 43; Mismatches 112; Indels 190; Gaps 19;

QY 1 MNRGVPRHLLVLTALALPAATQGNKVLGKKGDTVELTCTYASQKKSIOFMKNSNQIK 60
DB 19 MSRGVD-----IVL-----TQSPPTTAAISGPKVTITCRASSSVSYWYMYQOKS--- 62
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQG-NPFLIIKLNKIEDSDTYICEVEDQKEEYQ 119
DB 63 --GASPKLMIYDTSKLAAGVFNRFSGSGSTYSIAINMETEDATATYCO----- 111
QY 120 LLVFGLTANSDDLHLOGQSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDOS 179
DB 112 -----QWSTPLTF-----GSQTKLEIKRGGSGGSGGSGGSGQVQLKEA 152
QY 180 G-----TWCTVTLQNKQKVEFKIDIVLAFOKASSIYVKKEGEQVESEFPPLAFTV 229
DB 153 GPGIVQPTQTLSLTCTV-----SGFS----- 173
QY 230 EKLTSGSELWMQAEARASSKSWITFDLKNKEVSVKRVTDODPKLQMGKULPLHLTPQALP 289
DB 174 --LTSQGVH-----IRQPP-----GKGLW-----MGII 196
QY 290 QYAGSGNLTLEAKTKGLHGEVNLVYMRATOLQKNLTCEWGPSTPKMLSLKENKEA 349
DB 197 YDDGTDYNSAIKSR-----LSTSRDTS-----KSGVFLKINSLO- 231
QY 350 KVSKEKRPVWNLPEAGMOCQCL--LSDSGVLLSNIKVLPTWSTVPVCPAPEPSCDK 406
DB 232 -----TDDTMYTCARIHFDYWGQ-----GVMVTVSSDL-----EPKSCDK 267
QY 407 THTC-----PELLGSPVFLPFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDG 461
DB 268 THICPCPAPBELLGSPVFLPFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDG 327
QY 462 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGQ 521
DB 328 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGQ 387

```

```

QY 522 PREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDG 581
DB 388 PREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDG 447
QY 582 SFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 625
DB 448 SFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 491

RESULT 38
US-10-207-655-344
; Sequence 344, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207.655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 344
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion polypeptide
US-10-207-655-344

Query Match          36.8%; Score 1258; DB 14; Length 492;
Best Local Similarity 46.4%; Pred. No. 1.1e-80;
Matches 299; Conservative 42; Mismatches 113; Indels 190; Gaps 19;

```

```

QY 1 MNRGVPRHLLVLTALALPAATQGNKVLGKKGDTVELTCTYASQKKSIOFMKNSNQIK 60
DB 19 MSRGVD-----IVL-----TQSPPTTAAISGPKVTITCRASSSVSYWYMYQOKS--- 62
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQG-NPFLIIKLNKIEDSDTYICEVEDQKEEYQ 119
DB 63 --GASPKLMIYDTSKLAAGVFNRFSGSGSTYSIAINMETEDATATYCO----- 111
QY 120 LLVFGLTANSDDLHLOGQSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDOS 179
DB 112 -----QWSTPLTF-----GSQTKLEIKRGGSGGSGGSGGSGQVQLKEA 152
QY 180 G-----TWCTVTLQNKQKVEFKIDIVLAFOKASSIYVKKEGEQVESEFPPLAFTV 229
DB 153 GPGIVQPTQTLSLTCTV-----SGFS----- 173
QY 230 EKLTSGSELWMQAEARASSKSWITFDLKNKEVSVKRVTDODPKLQMGKULPLHLTPQALP 289
DB 174 --LTSQGVH-----IRQPP-----GKGLW-----MGII 196
QY 290 QYAGSGNLTLEAKTKGLHGEVNLVYMRATOLQKNLTCEWGPSTPKMLSLKENKEA 349
DB 197 YDDGTDYNSAIKSR-----LSTSRDTS-----KSGVFLKINSLO- 231
QY 350 KVSKEKRPVWNLPEAGMOCQCL--LSDSGVLLSNIKVLPTWSTVPVCPAPEPSCDK 406
DB 232 -----TDDTMYTCARIHFDYWGQ-----GVMVTVSSDL-----EPKSCDK 267
QY 407 THTC-----PELLGSPVFLPFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDG 461
DB 268 THICPCPAPBELLGSPVFLPFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDG 327
QY 462 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGQ 521
DB 328 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGQ 387
QY 522 PREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDG 581
DB 328 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGQ 387

```



Db 388 PREQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAWESNGQPENNYKTTTPVLDSDG 447  
Qy 582 SFFLYSKLTVDKSRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 625  
Db 448 SFFLYSKLTVDKSRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 491

RESULT 39  
US-09-773-877A-12  
; Sequence 12, Application US/09773877A  
; Publication No. US20030017977A1  
; GENERAL INFORMATION:  
; APPLICANT: Xia, Yu-Ping et al.  
; TITLE OF INVENTION: METHODS FOR TREATING INFLAMMATORY SKIN DISEASES  
; FILE REFERENCE: REG 710b  
; CURRENT APPLICATION NUMBER: US/09/773,877A  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 567  
; TYPE: PRF  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Flt(1-3)-Fc  
US-09-773-877A-12

Query Match 36.7%; Score 1254.5; DB 12; Length 567;  
Best Local Similarity 47.2%; Pred. No. 2.3e-80;  
Matches 298; Conservative 55; Mismatches 120; Indels 159; Gaps 24;

Qy 85 SLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQ-GGSLTTLT 143  
Db 3 SYMDTVGLCALSLCLITGSSSGSKLDPE-----LSLKGTHIMQAGOTLHL-- 51  
Qy 144 ESPPGSSPSVQCR-----SPRGKNIQGGK---TLVSQGLE 175  
Db 52 -----QCRGEAAHKMSLPEWVSKESERLSITSACGRN---GKQFCSTLTLLNTAQ 98  
Qy 176 LODSGTWTCTVL-----QNOKVEFKIDI-----VTLAFQKASSIVYKKEGQEVESF 223  
Db 99 ANHGFYSCRYKLAVPSTSKKETESAIYFISDTGRPFVEMWSEIPEIIMHTEGR--ELVI 156  
Qy 224 PLATVEKLTL-----GSGELMWQABRASSKSWITFDLKNKEVSRYATQD 269  
Db 157 PCRYTSPNITVTLTKKPFIDTLIPDGKRIW-----DSRKGFIISNATYKEIGL----- 204  
Qy 270 PKLQMGKKLPLHLTLPLQALPOYAGSGNLTLLAEKTKGL-HQEVNLV-----MRATQL 322  
Db 205 -----LTCENATV-----NGHL-----YKTNVLTNRQNTIITDVOISTPRPVKL 242  
Qy 323 QKN-----LTCENVGPTSPKMLSLKL--ENKEAKVSKR-----EKPV 358  
Db 243 LRGHITLVNCTATTPLNTRVQMTWSYPDEKKNKASVRRRIDQSNSHANIFYSVLTIDK-- 300  
Qy 359 WVLNPEAGMOCCLSDSGQVLLSESNIKVLPWTSTPVPAPAPKSCDKTHTC-----PEL 413  
Db 301 -MOKNDKGLTYCRVR-SGSPFSKSVNTSV-HIYDAGP---GEPKSCDKTHTCPCPAPBEL 354  
Qy 414 LGGSPVLFPPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFMWYVDGEVHNATKPREE 473  
Db 355 LGGSPVLFPPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFMWYVDGEVHNATKPREE 414  
Qy 474 QYNSTYRVVSVLTVLHODMNLNGKEYCKKVSNNKALPAPIEKTISAKAQPREPQVYTLPPS 533  
Db 415 QYNSTYRVVSVLTVLHODMNLNGKEYCKKVSNNKALPAPIEKTISAKAQPREPQVYTLPPS 474  
Qy 534 RDELTKNQVSLTCLVKGFYPSDIAWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDK 593  
Db 475 RDELTKNQVSLTCLVKGFYPSDIAWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDK 534  
Qy 594 SRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 625  
Db 594 SRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 625

Db 535 SRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 566

RESULT 40  
US-09-773-877A-20  
; Sequence 20, Application US/09773877A  
; Publication No. US20030017977A1  
; GENERAL INFORMATION:  
; APPLICANT: Xia, Yu-Ping et al.  
; TITLE OF INVENTION: METHODS FOR TREATING INFLAMMATORY SKIN DISEASES  
; FILE REFERENCE: REG 710b  
; CURRENT APPLICATION NUMBER: US/09/773,877A  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20  
; LENGTH: 567  
; TYPE: PRF  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Flt(1-3 R->N)-Fc (Mult4)  
US-09-773-877A-20

Query Match 36.7%; Score 1254.5; DB 12; Length 567;  
Best Local Similarity 47.2%; Pred. No. 2.3e-80;  
Matches 298; Conservative 55; Mismatches 120; Indels 159; Gaps 24;

Qy 85 SLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQ-GGSLTTLT 143  
Db 3 SYMDTVGLCALSLCLITGSSSGSKLDPE-----LSLKGTHIMQAGOTLHL-- 51  
Qy 144 ESPPGSSPSVQCR-----SPRGKNIQGGK---TLVSQGLE 175  
Db 52 -----QCRGEAAHKMSLPEWVSKESERLSITSACGRN---GKQFCSTLTLLNTAQ 98  
Qy 176 LODSGTWTCTVL-----QNOKVEFKIDI-----VTLAFQKASSIVYKKEGQEVESF 223  
Db 99 ANHGFYSCRYKLAVPSTSKKETESAIYFISDTGRPFVEMWSEIPEIIMHTEGR--ELVI 156  
Qy 224 PLATVEKLTL-----GSGELMWQABRASSKSWITFDLKNKEVSRYATQD 269  
Db 157 PCRYTSPNITVTLTKKPFIDTLIPDGKRIW-----DSRKGFIISNATYKEIGL----- 204  
Qy 270 PKLQMGKKLPLHLTLPLQALPOYAGSGNLTLLAEKTKGL-HQEVNLV-----MRATQL 322  
Db 205 -----LTCENATV-----NGHL-----YKTNVLTNRQNTIITDVOISTPRPVKL 242  
Qy 323 QKN-----LTCENVGPTSPKMLSLKL--ENKEAKVSKR-----EKPV 358  
Db 243 LRGHITLVNCTATTPLNTRVQMTWSYPDEKKNKASVRRRIDQSNSHANIFYSVLTIDK-- 300  
Qy 359 WVLNPEAGMOCCLSDSGQVLLSESNIKVLPWTSTPVPAPAPKSCDKTHTC-----PEL 413  
Db 301 -MOKNDKGLTYCRVR-SGSPFSKSVNTSV-HIYDAGP---GEPKSCDKTHTCPCPAPBEL 354  
Qy 414 LGGSPVLFPPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFMWYVDGEVHNATKPREE 473  
Db 355 LGGSPVLFPPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFMWYVDGEVHNATKPREE 414  
Qy 474 QYNSTYRVVSVLTVLHODMNLNGKEYCKKVSNNKALPAPIEKTISAKAQPREPQVYTLPPS 533  
Db 415 QYNSTYRVVSVLTVLHODMNLNGKEYCKKVSNNKALPAPIEKTISAKAQPREPQVYTLPPS 474  
Qy 534 RDELTKNQVSLTCLVKGFYPSDIAWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDK 593  
Db 475 RDELTKNQVSLTCLVKGFYPSDIAWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDK 534  
Qy 594 SRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 625  
Db 535 SRMOQGNVFCSCVMEALHNHYTQKSLSLSPG 566

RESULT 41

US-10-412-406-33  
; Sequence 33, Application US/10412406  
; Publication No. US20040058394A1  
; GENERAL INFORMATION:  
; APPLICANT: BIOGEN, INC.  
; APPLICANT: GABER, Ellen  
; APPLICANT: LYNE, Paul  
; APPLICANT: SALDHANA, Jose W.  
; TITLE OF INVENTION: HUMANIZED ANTI-LT-BETA-R ANTIBODIES  
; FILE REFERENCE: BINA100CN  
; CURRENT APPLICATION NUMBER: US/10/412,406  
; CURRENT FILING DATE: 2003-04-10  
; PRIOR APPLICATION NUMBER: 60/240,285  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/275,289  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/299,987  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: PCT/US01/32140  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 33  
; LENGTH: 4852  
; TYPE: PR1  
; ORGANISM: Homo Sapien  
US-10-412-406-33

Query Match 36.7%; Score 1253.5; DB 12; Length 4852;  
Best Local Similarity 45.5%; Pred. No. 4.8e-79;  
Matches 322; Conservative 53; Mismatches 180; Indels 153; Gaps 26;

QY 23 TQGNKVVLLGKGGTVELTCTASQKSIQFWKNSNQIKLIGNSGFLTKPSKLNDRADS 82  
DB 4193 TQSPSSLSASVGRVITTCAGODIKSYLSWYQOKP---GKAPKLIYATRLAUGVS 4248  
QY 83 RRLMDQG-NFPLIKKLIKEDSDTYC-----EVEDQKEVQLVFLGTA 127  
DB 4249 RFGSGSGSDTYITLTSLOPEDPATYICLOHGESPMTPGGGTLEIKRTYAASVF-IFP 4307  
QY 128 NSDTHLLOGQSLTLTLESPPGSSPSVQCR---SPRGKNIQ-----GKTLVSQEL 174  
DB 4308 PSDEQKSG-----TASVVCCLNNFYPRAKQWQVNAQSGNSQESVTEQ 4354  
QY 175 ELDDSGTWTCTVLOQKKEVF---KIDIVLAFQKASSIYK--KSGEYVESFPPLAFVY 229  
DB 4355 DSKDS--TYSLSSTYLTLSKADYEKHKYVACEVTHQGLSSPYTKSFNKECEVOL-----V 4408  
QY 230 EKLTG---SGELMWQERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPRLHLTP 285  
DB 4409 ESGGGLVKKPGGSL---RLSCAASGFTF---SDIYMYWFRQAP---GKGLEWATIS 4455  
QY 286 Q--ALPOYAGS--GNLTALAEATGKLGHOENVLVVMKATOL-----QKN--LTCEVWGP 333  
DB 4456 DGGSYTYPPDSVGRFTISHDNAKNSLYLQMSLRADETAIVVYCARBNENFYPDWQ 4515  
QY 334 TSPKMLSLKLENKAEKVPVWLNPEA-----GMQCLISD-----S 375  
DB 4516 GTTVTVSS-----ASTKGPSVFPPLAPSSKSTSGTALGLVVDYEPPEVTVSWS 4566  
QY 376 GOVLLSENIKVLTP-----TWSTVPQCAP-----EPK 402  
DB 4567 G--ALTSGVHTFPAVLOSGLYSLSSVTVYVSSSLGTQTYICNVNHPKSTKVDKVEPK 4624  
QY 403 SCDKTHTC-----PELLGSPSVFLPPKPKDTLMTSRPEVTCVVDVSHEDPEVKRMV 457  
DB 4625 SCCKTHCPCPAPPELLGSPSVFLPPKPKDTLMTSRPEVTCVVDVSHEDPEVKRMV 4684  
QY 458 VDGVEVNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISK 517  
DB 4685 VDGVEVNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISK 4744  
QY 518 AKQGRPREVYTLPPSRDELTKNQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTPPVLD 577

DB 4745 AKQGRPREVYTLPPSRDELTKNQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTPPVLD 4804  
QY 578 DSDGSFFLYSKLTVDKSRMGOQNVFSCVMEHALNHYTKSISLSPG 625  
DB 4805 DSDGSFFLYSKLTVDKSRMGOQNVFSCVMEHALNHYTKSISLSPG 4852

RESULT 42  
US-10-207-655-266  
; Sequence 266, Application US/10207655  
; Publication No. US20030118592A1  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Hayden-Ledbetter, Martha S.  
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
; FILE REFERENCE: 390069.401C1  
; CURRENT APPLICATION NUMBER: US/10/207,655  
; CURRENT FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 426  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 266  
; LENGTH: 552  
; TYPE: PR1  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion polypeptide  
US-10-207-655-266

Query Match 36.7%; Score 1252.5; DB 14; Length 552;  
Best Local Similarity 46.0%; Pred. No. 3e-80;  
Matches 297; Conservative 39; Mismatches 126; Indels 183; Gaps 18;

QY 1 MNRGVPFRLHLLVLOLALIPATQGNKVVLLGKGGTVELTCTASQKSIQFWK---KNSN 57  
DB 19 MNRGVD-----IVL-----TQSPATLSVTGPDVSLSCASQISIDYLMHYQOKSHE 65  
QY 58 QIKILGNQSGFLTKG--PSKLNDRADSRRLMDQGNFPLIKKLIKEDSDTYICEVEDQKE 116  
DB 66 SPRLIKTVASHISIGLPSRPSGSGS-----SDTLTINSVPEPDVGITYCQ----- 112  
QY 117 EVQLVFGGLTANSDFHLLLOGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQEL 176  
DB 113 -----HGHSFPWTF---GGGTLEIKRGGSGSGSGGSGGQIQ 150  
QY 177 QDSGTWCTVLOQKKEFKIDI VLAFOKASSIYKKEGEYVESFP--LAFT----- 228  
DB 151 VQSGP-----ELKKPGETVRLISCKASGYAFTTGMQ 181  
QY 229 -VEKLTSGELMWQERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPRLHLTPQ 287  
DB 182 WVOEMPGKLLK---IGWINTPLMSAKIC-----RRLQ----- 211  
QY 288 LPOYAGSNLTALAEATGKLGHOENVLVVMKATOLQKNLTCEVWGPSPKMLSLKLENK 347  
DB 212 -----GRFASLEISANTAVLQIS-----NLNDE-- 235  
QY 348 EAKVSRREKRVWVLNPEAGMOCCLSDSGQVLLSENIKVLPTW--STVPQCAPPEKSCD 405  
DB 236 -----DTATYFCVRSNGNRY---DLAFAYWGGQGLTVVSDLEPKSSD 275  
QY 406 KTHT-----CPELLGSPSVFLPPKPKDTLMTSRPEVTCVVDVSHEDPEVKRMVYVDG 460  
DB 276 KTHTSPSPAPPELLGSSSVFLPPKPKDTLMTSRPEVTCVVDVSHEDPEVKRMVYVDG 335  
QY 461 VEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKG 520  
DB 336 VEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKG 395  
QY 521 QPREPOVYTLPPSRDELTKNQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTPPVLDSD 580  
DB 396 QPREPOVYTLPPSRDELTKNQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTPPVLDSD 455



Db 625 SVLTVLHODWLNKCKVSKNKLPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQV 684  
QY 543 SLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPVLDSDGSFPLYSKLTVDKSRWQGQNVF 602  
Db 685 SLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPVLDSDGSFPLYSKLTVDKSRWQGQNVF 744  
QY 603 SCSVMEALHNHYTOKSLSPG 625  
Db 745 SCSVMEALHNHYTOKSLSPG 767

RESULT 45  
US-09-875-338-9  
Sequence 9, Application US/09875338  
Patent No. US20020095024A1  
GENERAL INFORMATION:  
APPLICANT: MIKESSELL, GLEN E.  
APPLICANT: CHANG, HAN  
APPLICANT: FINGER, JOSHUA N.  
APPLICANT: YANG, GUCHEN  
APPLICANT: LU, PIN  
APPLICANT: ZHOU, XIA-DI  
APPLICANT: PEACH, ROBERT  
TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR  
FILE REFERENCE: 3053-4071US2  
CURRENT APPLICATION NUMBER: US/09/875,338  
CURRENT FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: 60/272,107  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/209,811  
PRIOR FILING DATE: 2000-06-06  
NUMBER OF SEQ ID NOS: 94  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 698  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: fusion construct  
US-09-875-338-9

Query Match 36.6%; Score 1249; DB 9; Length 698;  
Best Local Similarity 43.8%; Pred. No. 7.3e-80;  
Matches 292; Conservative 56; Mismatches 121; Indels 198; Gaps 20;

QY 32 KKGDYVELTCTASQ-KKSIQFMKNSNQIKILGNQGSFLTGPCKLNDRADSRRLMDQ 90  
Db 156 RPDYTYITICSSYGYPEAEVFMQDQGVPLGN-----VTTSQMANEQ 200  
QY 91 NFPL-IINKLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLQOGSLTTLBSP- 147  
Db 201 LFDVHSILRVLVGANGTYSC-----LVARNVLOQDAH---SSVTITPORSPTG 245  
QY 148 -----GSSPSVQCR-SP-----RGKN--- 162  
Db 246 AVEVQPEDEPVVALVGTDAITLRCFSFSPGPSLAQLNLWQLTDTYKQLVHSFTBGRQGS 305  
QY 163 -----IOGKTLVSQLELDSDGTWCTTV-LONQKVEFKIDIIVILAFQKAS 208  
Db 306 AYANRTALFPDILAQGNASIRLQVRVADBSFTCFPSIRDPGSAVSLQVAPYSPSM 365  
QY 209 SIYVKE--GEQVESFPLAFVYEKLTG--SGELMWQAEERASSKSWITFDLKNKEVS 263  
Db 366 TLEBNKDLRPGDTV-----TITCSSYRGYPEAEVFMQD----- 398  
QY 264 KRVYQDPKLGKGLPLHLPLQALPQYAGSGNLTALAEKTKGLHGVNVLVYMRATQLQ 323  
Db 399 -----GQGVPL-----TGAVTTSQMANEQLF-DVHSVLRVVLGAN 433  
QY 324 KNLTCVWGPSTPLMLSLKLENKAKVSKREKPVWVLNPEAGMWQCLLSDSGVLLSEN 363

Db 434 GYVSLCLVNP-----VLQDA-----HGSVITITQ 458  
QY 384 IYVLTWSTVPDPCPADEPKSCDKTHTC-----PELLGSPVLFPPKPKDTLMISTPEV 438  
Db 459 PWTFP-----PEFEPEKSCDKTHTCPCPADELLGSPVLFPPKPKDTLMISTPEV 510  
QY 439 TCVVVDVSHEDPEVKNMVVDGVEVHNAKTKRREGQVNSTYVSVLTYLHODWLNKCKV 498  
Db 511 TCVVVDVSHEDPEVKNMVVDGVEVHNAKTKRREGQVNSTYVSVLTYLHODWLNKCKV 570  
QY 499 KCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPYPSDI 558  
Db 571 KCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPYPSDI 630  
QY 559 EWESNGQPENNYKTTTPVLDSDGSFPLYSKLTVDKSRWQGQNVFSCSVMEALHNHYTOK 618  
Db 631 EWESNGQPENNYKTTTPVLDSDGSFPLYSKLTVDKSRWQGQNVFSCSVMEALHNHYTOK 690  
QY 619 SLSPG 625  
Db 691 SLSPG 697

RESULT 46  
US-10-077-023-9  
Sequence 9, Application US/10077023  
Publication No. US20030031675A1  
GENERAL INFORMATION:  
APPLICANT: MIKESSELL, GLEN E.  
APPLICANT: CHANG, HAN  
APPLICANT: FINGER, JOSHUA N.  
APPLICANT: YANG, GUCHEN  
APPLICANT: LU, PIN  
APPLICANT: ZHOU, XIA-DI  
APPLICANT: PEACH, ROBERT  
TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR  
FILE REFERENCE: 3053-4071US3  
CURRENT APPLICATION NUMBER: US/10/077,023  
CURRENT FILING DATE: 2002-02-15  
PRIOR APPLICATION NUMBER: 60/272,107  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/209,811  
PRIOR FILING DATE: 2000-06-06  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 698  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: fusion construct  
US-10-077-023-9

Query Match 36.6%; Score 1249; DB 14; Length 698;  
Best Local Similarity 43.8%; Pred. No. 7.3e-80;  
Matches 292; Conservative 56; Mismatches 121; Indels 198; Gaps 20;

QY 32 KKGDYVELTCTASQ-KKSIQFMKNSNQIKILGNQGSFLTGPCKLNDRADSRRLMDQ 90  
Db 156 RPDYTYITICSSYGYPEAEVFMQDQGVPLGN-----VTTSQMANEQ 200  
QY 91 NFPL-IINKLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLQOGSLTTLBSP- 147  
Db 201 LFDVHSILRVLVGANGTYSC-----LVARNVLOQDAH---SSVTITPORSPTG 245  
QY 148 -----GSSPSVQCR-SP-----RGKN--- 162  
Db 246 AVEVQPEDEPVVALVGTDAITLRCFSFSPGPSLAQLNLWQLTDTYKQLVHSFTBGRQGS 305  
QY 163 -----IOGKTLVSQLELDSDGTWCTTV-LONQKVEFKIDIIVILAFQKAS 208

```

Db 306 AYAARIALFPDLAQNASLRLOVRVADSGSFTCFVSIRDFGSAANSLQVAAPSKPSM 365
Qy 209 SIIVKKE---GEQYEFSPFLAFTVEKLTG--SGELMWOAERASSKSMITFDLKNKEVS 263
Db 366 TLEPNKDLRPDQTV-----TTCSSYRGYPDAEFVFD----- 398
Qy 264 KRVTPDPLQNGKTLPLHLTLPOALPOYAGSGNLTLLAEAKTGKLDHENVLVNRAATOLQ 323
Db 399 -----GQGVPL-----TGNTTSGMANEGQLF-DVHSLVLAIVGAN 433
Qy 324 KNLTCGEWGPFSPLKMLSLKLENKEAKYSKREKPVVNLNPAQMMOCLLSDSGVLLASN 383
Db 434 GTVSLVARNP-----VLQODA-----HGSVTLTGQ 458
Qy 384 IKVLPTMSTPVPAPAPBPKCDKHTC-----PELLGSPSVFLFPKPKDTLMSRPPEV 438
Db 459 PMTP-----PEPEKSCDKHTCPCPAPPELLGSPSVFLFPKPKDTLMSRPPEV 510
Qy 439 TCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYRVVSVLTVLDHQMNGKEY 498
Db 511 TCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYRVVSVLTVLDHQMNGKEY 570
Qy 499 KCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 558
Db 571 KCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 630
Qy 559 EWESNGQPENNYKTTTPVLDSGSEFLYSKLTVDKSRKQGNVSCSMHEALNHYTQK 618
Db 631 EWESNGQPENNYKTTTPVLDSGSEFLYSKLTVDKSRKQGNVSCSMHEALNHYTQK 690
Qy 619 SLSLSPG 625
Db 691 SLSLSPG 697

```

```

RESULT 47
US-09-773-877A-14
; Sequence 14, Application US/09773877A
; Publication NO. US20030017977A1
; GENERAL INFORMATION:
; APPLICANT: Xia, Yu-Ping et al.
; TITLE OF INVENTION: METHODS FOR TREATING INFLAMMATORY SKIN DISEASES
; FILE REFERENCE: REG 710b
; CURRENT APPLICATION NUMBER: US/09/773,877A
; CURRENT FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Flt1(1-3 deltab)-Fc (Mut1)
US-09-773-877A-14

```

```

Query Match 36.5%; Score 1245.5; DB 12; Length 557;
Best Local Similarity 47.2%; Pred. No. 9.6e-80;
Matches 295; Conservative 54; Mismatches 121; Indels 155; Gaps 24;

Qy 85 SLMDQGNPLIIRKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDDLHQQGSLTTL 143
Db 3 SYMDTGVLCLALLSCLLITGSSSGSKLDPF-----LSLKGTHIMQAGTIIH-- 51
Qy 144 ESPGSSPSVQCR-----SPRGKNIQCKR---TLVSQLE 175
Db 52 -----QCRGAHAKMSLPKEMVSKESRLSITKACGRN---GKQFCSTLTLLNQAQ 98
Qy 176 LODSGTWTCTVL-----QNOKVEPKIDI-----VTLAQKASSIYYKKEGQVESP 223
Db 99 ANHGGFYSCKYLAVPTSKKETESAIYIFISDTGRPFPEMSEIPEIIMHTEGR--ELVI 156
Qy 224 PLAFVETLT-----GSGELMWOAERASSKSMITFDLKNKEVSVAIVTQD 269

```

```

Db 157 PCRVTSNITVTLLKPKPLDILLDPGRIIWM-----DSRKGIISNATYKEIGL----- 204
Qy 270 PKIQMGKKLPLHLTLPOALPOYAGSGNLTLLAEAKTGKL-HQENVLVN-----MRAATOL 322
Db 205 -----LTCERTV-----NGHL-----YKINYLTHRQNTIIDIQISPRPVKL 242
Qy 323 QKN-----LTCEWGPFSPLKMLSLKLENKEAK-----VSKREKPVVNLNPEA 365
Db 243 LRHGTILVNCATATPLNTRVQMTWSYPDELQDSHANIFYSVLTIDKMQ-----NKDK 296
Qy 366 GMMOCLLSDSGQVLLASNIKYLPMTSTPVPAPAPBPKCDKHTC-----PELLGSPSVF 420
Db 297 GLYTCRVYR-SGPSEKSVNTSV-HIYDRAKP---GEPEKCDKHTCPCPAPPELLGSPSVF 351
Qy 421 LFPKPKDTLMSITSTPPTCVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYR 480
Db 352 LFPKPKDTLMSITSTPPTCVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYR 411
Qy 481 VVSVLTVDHQMNGKEYKCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKN 540
Db 412 VVSVLTVDHQMNGKEYKCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKN 471
Qy 541 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLYSKLTVDKSRKQGN 600
Db 472 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLYSKLTVDKSRKQGN 531
Qy 601 VFSGSVNHEALNHYTQKSLSLSPG 625
Db 532 VFSGSVNHEALNHYTQKSLSLSPG 556

```

```

RESULT 48
US-10-207-655-240
; Sequence 240, Application US/10207655
; Publication NO. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 240
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion polypeptide
US-10-207-655-240

```

```

Query Match 36.4%; Score 1243; DB 14; Length 500;
Best Local Similarity 47.8%; Pred. No. 1.2e-79;
Matches 295; Conservative 37; Mismatches 127; Indels 158; Gaps 18;

Qy 23 TQGNKVLKGGKDTVBLICTASQKSIQIFHWKNSNOIKIIGNQGSLTKGSPSKLNDRA 82
Db 27 SQSPAILASBPGEKVTWTCRASSSVS--YHMYQOKP-----GSSPKPVIYAPSNLASGVPA 81
Qy 83 RRLIMDQG-NPPLIKLIKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDDLHQQGSLT 141
Db 82 RFSGSGGTYSILTSISVEADATYYCQ-----QMSFNPPTF---GAGTKL 125
Qy 142 TLESPPGSSPSVQCRSPRGKNIQCKTLVSQLELODGTWTCTVLQNOKVEPKIDIIV 201
Db 126 ELKQGGSG-----GGGSGGGGS--SQAVIQSGA-----ELV- 156
Qy 202 LAFOKASSIYYKKEGQVESPFLAFTVEKLTG--SGELMWOAERASSKSMITFDL 257
Db 157 -----RSGASVWKSC-----KASGTFPTSYMMHWKQTPROGLEWIG---- 193
Qy 258 NKEVSVAIVTQDPLQNGKTLPLHLTLPOALPOYAGSGNLTLLAEAKTGKLDHENVLVN 317

```

```

Db 194 -----ATTPNGDPTSYNQK-RK-----211
QY 318 RATQLOKUNLTCCEVWGPTSPKMLSLKLENKAVSKREKRVWVWLNPEAGMOCCLSDSGQ 377
QY 212 -----ATLTVDKSSSTAYWQJSLTSE-----DSAVYFC-----AR 242
Db 378 VLLESN-----IKVLPWTSTPVCPCAPBPCKCDKTHTC-----PELLGGSVFLFPKPKD 428
QY 243 VVYYSNSYWFDMWGTGTTVTWSSDOEPKSCDKTHTCPPCPAPELLGGSVFLFPKPKD 302
Db 429 TLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENHNAKTKEPREQYNSTYRVSVLTCLV 488
QY 303 TLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENHNAKTKEPREQYNSTYRVSVLTCLV 362
Db 489 HODMLNGKEYKCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDELTKNQVSLTCLV 548
QY 363 HODMLNGKEYKCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDELTKNQVSLTCLV 422
Db 549 KGFYPSDIAVWESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCSSVMH 608
QY 423 KGFYPSDIAVWESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCSSVMH 482
Db 609 EALHNHYTOKSLSLSPG 625
QY 483 EALHNHYTOKSLSLSPG 499
Db

RESULT 49
US-10-207-655-398
; Sequence 398, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069, 401C1
; CURRENT APPLICATION NUMBER: US/10/207, 655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 398
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion polypeptide
US-10-207-655-398

```

```

Query Match 36.4%; Score 1243; DB 14; Length 500;
Best Local Similarity 47.8%; Pred. No. 1,2e-79;
Matches 295; Conservative 37; Mismatches 127; Indels 158; Gaps 18;

QY 23 TQGNKVVIGKGGPVELTCTASQKKSIOFMKNSNQIKILGNQGSFLTKGSPSKLNRADS 82
Db 27 SQSPALISAPGKVTMTCPAASSVS-YMHWYQKP-----GSSPKPIYAPSNLASGVRA 81
QY 83 RSLMDQG-NFPIIKNLKIEDSYICVEVDQKEVQLLVFGLTANSDTHLLQGOSLTL 141
Db 82 RFGSGSGGTSLTISRVEAEDATATYCYQ-----QMSFNPPTF---GAGTKL 125
QY 142 TLSSPSSPSVQCRSRGNIGQKTLSTQJLEODSGTWTCTVLOKQKVEFKIDIV 201
Db 126 ELDDGGSG-----GGSGSGGGS---SQAYLQSGA-----ELV- 156
QY 202 LAFQKASSIYKKEGEVSEFPLAFVEKLTG-----SGELMMQAEPRASSKSMITFDLK 257
Db 157 -----RPGASVKKSC-----KASGTTFTSYMMHWVKQTPROGLEWIG----- 193
QY 258 NKEVSVKRVTDQPKLQWKKLPLHLTLPOALPOYAGSGNLTLALNAKTKLHOEVNLVVM 317
Db 194 -----ATYPGNGDTSYNQKFK-GK-----211

```

```

QY 318 RATQLOKUNLTCCEVWGPTSPKMLSLKLENKAVSKREKRVWVWLNPEAGMOCCLSDSGQ 377
Db 212 -----ATLTVDKSSSTAYWQJSLTSE-----DSAVYFC-----AR 242
QY 378 VLLESN-----IKVLPWTSTPVCPCAPBPCKCDKTHTC-----PELLGGSVFLFPKPKD 428
Db 243 VVYYSNSYWFDMWGTGTTVTWSSDOEPKSCDKTHTCPPCPAPELLGGSVFLFPKPKD 302
QY 429 TLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENHNAKTKEPREQYNSTYRVSVLTCLV 488
Db 303 TLMISRTPEVTCVVDVSHEDPEVKFNWYDGVENHNAKTKEPREQYNSTYRVSVLTCLV 362
QY 489 HODMLNGKEYKCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDELTKNQVSLTCLV 548
Db 363 HODMLNGKEYKCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDELTKNQVSLTCLV 422
QY 549 KGFYPSDIAVWESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCSSVMH 608
Db 423 KGFYPSDIAVWESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNVFCSSVMH 482
QY 609 EALHNHYTOKSLSLSPG 625
Db 483 EALHNHYTOKSLSLSPG 499

```

```

RESULT 50
US-10-683-255-2
; Sequence 2, Application US/10683255
; Publication No. US20040063910A1
; GENERAL INFORMATION:
; APPLICANT: Kavanaugh, William M.
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR
; TITLE OF INVENTION: RECEPTOR-IMMUNOGLOBULIN FUSION
; FILE REFERENCE: PP01474, 101
; CURRENT APPLICATION NUMBER: US/10/683, 255
; CURRENT FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 09/499, 846
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/119, 002
; PRIOR FILING DATE: 1999-02-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-683-255-2

```

```

Query Match 36.4%; Score 1243; DB 12; Length 622;
Best Local Similarity 48.3%; Pred. No. 1,7e-79;
Matches 299; Conservative 47; Mismatches 119; Indels 154; Gaps 23;

QY 109 CEVEDQKEVQLLVFGL-TANSDTHLLQGOSLTLTLSSPGSSPSVQC--RSPRGK----- 161
Db 55 CLRBDVQGINMLRDCVQLAESNRRTITGEVEVQ-DVFPADSGLACTYSSSGSDTTY 113
QY 162 ---NIQGGKTLVSQJLEODSGT-----WTCTVLOKQKVEFKIDIV 202
Db 114 FSNVNSDALPSSSEDDDDSSSEKETDNTKPNPAPYWT-----SPKMEKKLHAV-- 166
QY 203 AQQKASSIYKKEGEVSEFPLAFVE-KLQSGSLMMQAEPRASSKSMITFDLKN-KE 260
Db 167 -----PAAKTVKPKCPSSG-----TPNPTLRM-----LNQKKE 194
QY 261 VSVKRVTDQPKLQWKK-KLPLHLTLPOALPOYAGSGNLTLALNAKTKLHOEVNL-V 315
Db 195 FK-----PDHRIGYKRYATWISIMSVF--SDKNYTCIYVENHGSIMHTYQDLY 245
QY 316 VWRATQ--LQKNL-----TCEVWGPTSPKMLSLKLE----- 345
Db 246 VERSPHRPILQGLPANKTVLALGSNVFMCVKYSDQPHIOMLKHIEVNSKIGPDNLPY 305

```

[illegible]

## RESULT 51

```

US-10-232-838-18
; Sequence 18, Application US/10232838
; Publication No. US20030064053A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Shengliang
; APPLICANT: Martin, Jean-Francois
; APPLICANT: Liu, Dayou
; TITLE OF INVENTION: MULTIVALENT PROTEIN CONJUGATE WITH MULTIPLE LIGAND-BINDING DOMAIN
; TITLE OF INVENTION: RECEPTORS
; FILE REFERENCE: 26050-707
; CURRENT APPLICATION NUMBER: US/10/232,838
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US 06/316,718
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 949
; TYPE: PR1
; ORGANSIM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: MVP-C
; US-10-232-838-18

```

Query Match	36.4%	Score 1243	DB 12	Length 949
Best Local Similarity	38.1%	Pred. No. 3e-79		
Matches	327	Conservative	66	Mismatches 150; Indels 316; Gaps 31
Qy	33	KGDIIVELTCTASQ--KSIQFHWNKSNQIKILGNQGSFLTKGPSKLNDRADRSRLMDQG	90	
Db	140	RGHTLVLMCTATTPLNTRVQMTWSY-----PDKNRASVRRI--DOS	181	
Qy	91	N-----FPLIKLKIEDSDPTYCEVEDQKEVQLVFGLTANSDPHLQGGSLTTLT	143	
Db	182	NSHANIFVSVLTIDMQMKDKGLYTCRRSGS-----FKSVTTSHTIIDKAFITYKH	234	
Qy	144	ESPPSSPSVOCRSFRGNIOGKTL---SVSQLELQDSGTWTCTL--	187	
Db	235	GAGGGSGSG---GSGSGGGEVGGAMDLLIINININSPLVSDAETSFLTCLASGWRPHEPITI	291	
Qy	188	-----ONQKVEFKIDIVLVAFOKASSIVKKEGEQVEFSPFLAFVETKLS---	235	
Db	292	GRDFEALNNQNDPLEVTQDV---TRBAKKVYVWIRE-----KSKINGAIFCC	336	
Qy	236	-GELTMQW-----ERASSSKSWITFDL--KMKSVYKRV---TQD-----	269	
Db	337	EGRVNGEIRIRITWKMQRQOASFPLPTLLTWTYDQKGNVMSFKKVLKEEDAVIYKNGSFI	396	

OY	270	----	PKLQMGKJRLHNTLQALPOVAG-----	SGNLTLL-----	LEAK-----	3040
D5	397	HSVPRHSVPDILEVH--LPHAPQDQVYSARYIGNLTFTAFTRLLYRCEAKMGPEC				454
OY	305	----	TGKLHQEVLNVLM-----	RATOLO-----		3230
D5	455	NHLCTACMNNGVCHDPTGSCICRPPGMGRTECKACALHTTPTRTCKERSGQEGCKSVFC				514
OY	324	----	KNLTCE-----	VMGP-----		333
D5	515	LPPDYGCSCATGWKGLQCNBEACHPGFYGDCKLRCSCNNGBMCDRFOGCLSPGMQGLQC				574
OY	334	----	TSPKLM--LSLKL-----	NEARVYSREKRYM-----		355
D5	575	EREGIPRMPTKIVDLDPDHLLEVNSGKPNPICASGMPLPTNEEMTLVXPDGTVLHPKQFNH				634
OY	360	----	VLNPEAGMMQCLLSD--SGVLLLESNI--	KVLPTMSTVPKPCAP----		399
D5	635	TDHPSVALFTIRHLPLPDGSGVAVCSVNYVAGVMEKPFNISVXVLPK-----	PLNAPVID			689
OY	400	----	EPNSCDKTHTC-----	PELLGSPSVLFPKPK		428
D5	690	TGHNFAVINISSEPYFGDGPRIKSKKLVDNBSKSCDHTHCPCPCAPABELLGGSPVLFPPKP				749
OY	427	KDPLTMSRPELVYCVVVDVSHEDPEVKFMVYVDGVEVHNATKPREQVNSTYVSVVLT				488
D5	750	KDPLTMSRPELVYCVVVDVSHEDPEVKFMVYVDGVEVHNATKPREQVNSTYVSVVLT				809
OY	487	VLHQDMLNGKEYCKVSNKALPAPLEKTIISAKAQGPPEQVYTLTPPSRDELTKNQVSLTC				546
D5	810	VLHQDMLNGKEYCKVSNKALPAPLEKTIISAKAQGPPEQVYTLTPPSRDELTKNQVSLTC				865
OY	547	LVKGFYPSDIAVEMESNGQPENNYKTTTPPVLDSGSPFLYSKLTVDXSRWQGNVPFSCSV				606
D5	870	LVKGFYPSDIAVEMESNGQPENNYKTTTPPVLDSGSPFLYSKLTVDXSRWQGNVPFSCSV				929
OY	607	MHEALNHNHTQKSLSLSPG	625			
D5	930	MHEALNHNHTQKSLSLSPG	948			

## RESULT 52

US-09-875-338-5  
Sequence 5: Application US/09875338  
Patent No. US20020095024A1  
GENERAL INFORMATION:  
APPLICANT: MIKESELL, GLEN E.  
APPLICANT: CHANG, HAN  
APPLICANT: FINGER, JOSHUA N.  
APPLICANT: YANG, GUICHEN  
APPLICANT: LU, PIN  
APPLICANT: ZHOU, XIA-DI  
APPLICANT: PEACH, ROBERT  
TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR  
TITLE OF INVENTION: IMMUNOMODULATION  
FILE REFERENCE: 3053-4071US2  
CURRENT APPLICATION NUMBER: US/09/875.338  
CURRENT FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: 60/272,107  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/209,811  
PRIOR FILING DATE: 2000-06-06  
NUMBER OF SEQ ID NOS: 94  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 480  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Fusion Construct  
US-09-875-338-5





Db 82 RPSGSGSTSYSLTISRVEADATYYCQ-----QMSFNPPTF---GAGTKL 125  
Qy 142 TLSPGSSPSVQCRSPRKNIOGGKTLVSQLELDQSGTWTCTYLQKQKVEFKIDIV 201  
Db 126 ELKGGGSG-----GGSGGGGS---SQAYLQSGA-----ELV- 156  
Qy 202 LAFKASSIYKKEGEQVEFSPLATFVEKLTG---SGELMWAERASSKSMITPDLK 257  
Db 157 -----RPGASVKMSC-----KASGYFTSYNNHWYKQTPROGLEMIIG--- 193  
Qy 258 NKEVSVKRVTPDPLQNGKCLPLHLTLPLQALPOYAGSGNLTLEAKTGKLGHEVNLVVM 317  
Db 194 -----ALYPGNDITSYNOKFK-GK----- 211  
Qy 318 RATOLQKNTLCEWNGPTSPKMLSLKLENKE---AKYSKREKVPWVLNPEAGMGCCLS 373  
Db 212 -----ATLTVDKSSSTAYMQSLTSEDSAVYFCARVYVNSGYTFD----- 254  
Qy 374 DSGVLLSESNIKVLPTW--STPVPCEPAPRPSCKTKHTC-----PELLGSPSVLFPKP 426  
Db 255 -----VMGTGTTVYSDQEPKSCDKHTCPPCPAPPELLGSPSVLFPKP 299  
Qy 427 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRREQYNSYTRVSVLT 486  
Db 300 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRREQYNSYTRVSVLT 359  
Qy 487 VLHODWLNGEKYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTLC 546  
Db 360 VLHODWLNGEKYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTLC 419  
Qy 547 LVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDSKRWQOGNVFSCSV 606  
Db 420 LVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDSKRWQOGNVFSCSV 479  
Qy 607 MHEALHNHYTQKSLSLSPG 625  
Db 480 MHEALHNHYTQKSLSLSPG 498

RESULT 55  
US-10-207-655-148  
Sequence 148, Application US/10207655  
Publication No. US20030118592A1  
GENERAL INFORMATION:  
APPLICANT: Ledbetter, Jeffrey A.  
TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
FILE REFERENCE: 390069.401C1  
CURRENT APPLICATION NUMBER: US/10/207,655  
CURRENT FILING DATE: 2002-07-25  
NUMBER OF SEQ ID NOS: 426  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 148  
LENGTH: 499  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Mouse-Human hybrid fusion protein  
US-10-207-655-148

Query Match 36.4%; Score 1242.5; DB 14; Length 499;  
Best Local Similarity 47.5%; Pred. No. 1.3e-79;  
Matches 294; Conservative 33; Mismatches 129; Indels 163; Gaps 18;

Qy 23 TQGNKVVLGKKGDVVELTCTASQKSIQFHWKNSNOIKLGNQGSFLTGPSKLNDRADS 82  
Db 27 SQSPALISAPGKVTMTCTASSSVS-YMHWYQCKP-----GSSPKPMIYAPSNLASGVA 81  
Qy 83 RRLSLMDQG-NPPLIIKLIKIDSDTYICEVEDQKEVQLVFGLTANSQTHLLQGSGLTL 141  
Db 82 RPSGSGSTSYSLTISRVEADATYYCQ-----QMSFNPPTF---GAGTKL 125  
Qy 142 TLSPGSSPSVQCRSPRKNIOGGKTLVSQLELDQSGTWTCTYLQKQKVEFKIDIV 201

Db 126 ELKGGGSG-----GGSGGGGS---SQAYLQSGA-----ELV- 156  
Qy 202 LAFKASSIYKKEGEQVEFSPLATFVEKLTG---SGELMWAERASSKSMITPDLK 257  
Db 157 -----RPGASVKMSC-----KASGYFTSYNNHWYKQTPROGLEMIIG--- 193  
Qy 258 NKEVSVKRVTPDPLQNGKCLPLHLTLPLQALPOYAGSGNLTLEAKTGKLGHEVNLVVM 317  
Db 194 -----ALYPGNDITSYNOKFK-GK----- 211  
Qy 318 RATOLQKNTLCEWNGPTSPKMLSLKLENKE---AKYSKREKVPWVLNPEAGMGCCLS 373  
Db 212 -----ATLTVDKSSSTAYMQSLTSEDSAVYFCARVYVNSGYTFD----- 254  
Qy 374 DSGVLLSESNIKVLPTW--STPVPCEPAPRPSCKTKHTC-----PELLGSPSVLFPKP 426  
Db 255 -----VMGTGTTVYSDQEPKSCDKHTCPPCPAPPELLGSPSVLFPKP 299  
Qy 427 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRREQYNSYTRVSVLT 486  
Db 300 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRREQYNSYTRVSVLT 359  
Qy 487 VLHODWLNGEKYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTLC 546  
Db 360 VLHODWLNGEKYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTLC 419  
Qy 547 LVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDSKRWQOGNVFSCSV 606  
Db 420 LVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDSKRWQOGNVFSCSV 479  
Qy 607 MHEALHNHYTQKSLSLSPG 625  
Db 480 MHEALHNHYTQKSLSLSPG 498

RESULT 56  
US-10-053-530-15  
Sequence 15, Application US/10053530  
Publication No. US2003013939A1  
GENERAL INFORMATION:  
APPLICANT: Ledbetter, Jeffrey  
TITLE OF INVENTION: Binding Domain-Immunoglobulin Fusion Proteins  
FILE REFERENCE: 390069.401  
CURRENT APPLICATION NUMBER: US/10/053,530  
CURRENT FILING DATE: 2002-01-17  
PRIOR APPLICATION NUMBER: US 09/765,208  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 15  
LENGTH: 499  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: MOUSE-HUMAN HYBRID FUSION PROTEIN  
NAME/KEY: SITE  
LOCATION: (1)..(265)  
OTHER INFORMATION: MOUSE ANTI-HUMAN CD20 SCFV: 2H7  
NAME/KEY: DOMAIN  
LOCATION: (266)..(499)  
OTHER INFORMATION: HUMAN IGG1 WILD TYPE HINGE, CH2, CH3 FC  
US-10-053-530-15

Query Match 36.4%; Score 1242.5; DB 14; Length 499;  
Best Local Similarity 47.5%; Pred. No. 1.3e-79;  
Matches 294; Conservative 33; Mismatches 129; Indels 163; Gaps 18;

Qy 23 TQGNKVVLGKKGDVVELTCTASQKSIQFHWKNSNOIKLGNQGSFLTGPSKLNDRADS 82  
Db 27 SQSPALISAPGKVTMTCTASSSVS-YMHWYQCKP-----GSSPKPMIYAPSNLASGVA 81

```

Query Match 36.4%; Score 1242; DB 16; Length 634;
Best Local Similarity 45.5%; Pred. No. 2e-79;
Matches 298; Conservative 39; Mismatches 122; Indels 196; Gaps 18;

QY      10 LLLVLTAL--LP-----PATOQNKVYLGKGGDTVELTCTAASQKSIQPHM---K 54
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       2 LLLVTSLLCELPHPAFLLIPDIQWQTSSLSASLGVRVITSCRAQSDISKYLNWYQK 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      55 NSNQIKLIGNQSPITKG-PKSLNDRADRSKSLMDQGNFLLIKLKLKEDSDTYICEVD 113
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

```

RESULT 58
US-09-875-338-17
; Sequence 17, Application US/09875338
; Patent No. US20020095024A1
; GENERAL INFORMATION:
; APPLICANT: MIKESELL, GLEN E.
; APPLICANT: CHANG, HAN
; APPLICANT: FINGER, JOSHUA N.
; APPLICANT: YANG, GUICHEN
; APPLICANT: LU, PIN
; APPLICANT: ZHOU, XIA-DI
; APPLICANT: PEACH, ROBERT
; TITLE OF INVENTION: B7-RELATED NUCLEIC ACIDS AND POLYPEPTIDES USEFUL FOR
; FILE REFERENCE: 3053-4071US2
; CURRENT APPLICATION NUMBER: US/09/875,338
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/272,107
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/209,811
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: fusion construct
US-09-875-338-17
Query Match 36.4%; Score 1241.5; DB 9; Length 451;
Best Local Similarity 50.0%; Freq. No. 1.4e-79;

```



```

Query Match      36.3%; Score 1239.5; DB 12; Length 555;
Beat Local Similarity 45.1%; Pred. No. 2.5e-79; Indels 139; Gaps 18;
Matches 296; Conservative 55; Mismatches 167;

Qy      4 GVFPRHLLLVLOALPPAATOG-----NKVVLGKKDITVELTCTASQKKSIO 50
Db      2 GAPASALLLLLLLFPACCMARPGANLSQDSQSPMTSDPTVA--GGTVLVCQYKDHDDSS 59
Qy      51 FHWKNSNQIKIL-----GNQSFITKGPSTKLNDRADSRSLMDQGNPPLIIKLIKIE 102
Db      60 LQWNSPAAQQTLYEGEKKALNDNRIOALTSTPRHEIS-----ISISVALA 103
Qy      103 DSDTYICE-----VEDQKEVQLVFF--GLTNSDTHLLQGGSLTTLTSPGSSPSVQ 154
Db      104 DEGEYTSITMPRTAKSLVTVLGIPQKPIITGYKSSLEKOTATLNCPS--SGSKPAAR 162
Qy      155 CRSPRGKNIQGGKTLVSYQLELDQSGTWTCTVLQNKQKVEFKIDIVLAFQKASSIYKK 214
Db      163 LTRKRGQELHGEETRIQ---EDPNKTKTV-----SSSVTFQV 198
Qy      215 EGQGVFSPFLAFTVEKLTGSGGELMQAERASSSKSNITPDLKQKEVSVKRVTD--PKIQ 273
Db      199 TRDDGASIVCSVNHEGLK-----ADRSQRIEVLV-----TPRAMIRPDPHPR 245
Qy      274 MGKGLPLHLPLPALPOLYASSGNLTALAEKTKLQHEVNLVVMKRAQOLOKILCEVWGP 333
Db      246 EGQGLLLHC-----EGRGN-----PVPQOYLWEXGS 272
Qy      334 TSPKLMLSLTLENKAKVSKREKRVWVNLDEAGWMOCLLSDSGVLLDSNRIKULPTWSTP 393
Db      273 VPP-----LMTQGESALIFP-----FLKSDSGTGTATSN-----MSGYKAYITLVNN 317
Qy      394 VPCDAPBPKSCDKTHTC-----DELLGSPSVFLPPPKPKDTLMISRTPEVTCVVVDVSH 448
Db      318 DPSPVPSGSRDKHTCTPCPAPPELLGSPSVFLPPPKPKDTLMISRTPEVTCVVVDVSH 377
Qy      449 DPEYKFWIYDGVAVHNAKTKPREEOYNSTYRVVSVLTTLHOMLNKEYEKCKVSNKALP 508
Db      378 DPEYKFWIYDGVAVHNAKTKPREEOYNSTYRVVSVLTTLHOMLNKEYEKCKVSNKALP 437
Qy      509 APIEKTISKAGOREPOVYTLRPSRDELTKNOVSLCLYKGFPSDIAVEMESNGOPEN 568
Db      438 VPIEKTISKAGOREPOVYTLRPSRDELTKNOVSLCLYKGFPSDIAVEMESNGOPEN 497
Qy      569 NYKTTPEVLDSGDSFFLYSKLTVDKSKRMQOQNVFSCVMEHALNHNHTOKSLSPG 625
Db      498 NYKTTPEVLDSGDSFFLYSKLTVDKSKRMQOQNVFSCVMEHALNHNHTOKSLSPG 554

RESULT 62
US-10-120-198B-2
: Sequence 2, Application US/10120198B
: Publication No. US20030215427A1
: GENERAL INFORMATION:
: APPLICANT: Jensen, Michael
: TITLE OF INVENTION: CE7-SPECIFIC REDIRECTED IMMUNE CELLS
: FILE REFERENCE: 1954-337
: CURRENT APPLICATION NUMBER: US/10/120,198B
: PRIOR FILING DATE: 2002-04-11
: PRIOR APPLICATION NUMBER: 60/282,859
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 631
: TYPE: PRT
: ORGANISM: artificial sequence
: FEATURE:
: OTHER INFORMATION: mouse-human chimera
US-10-120-198B-2

```

Best Local Similarity 45.6%; Pred. No. 3e-79;  
Matches 295; Conservative 47; Mismatches 116; Indels 189; Gaps 20;

```

OY 5 VPFRLILVLALALPAATOGKIVLGGKGDVELTCTASOKKSIQF--HM---KNSQI 59
DB 13 LHPAFALIPQOQOPAE-----LVKPGASVKLSKASGFTTGMMHWKQPGGL 67
OY 60 KILNQSFLTKPSKNDRADSRRLW-DQGNFP--LIINKLKIEDDTYLCEVEDQKE 116
DB 68 EMIGINP--SNGRRNVERFRSKATLVDSSTTAFAQGLTSEDAVFCARD----- 121
OY 117 EVQLLVFGLTANSDTHLQGGSLITLESPPSSPSVOCRSPRGKNIQGGKTLVSQ--- 173
DB 122 ---YGTSYNFD---YMGQGTTLTVSSGGGGSGG-----GGSGGGSDIQWTQSSS 166
OY 174 ---LELDQSGTMTCTVLONOKKVEPKIDIVILAFQKASSIYKKEGEVEFSFLATVE 230
DB 167 SPSVSLGDRVITTC-----KANEIDI-----186
OY 231 KLTSGLLMOAERASSSKSWITFDLKNKEVSVKKVTQDPKLQMGKLLPLHLTPQALPQ 290
DB 187 ---NNRLAWYQOTFGNS-----PRLISGATNLVTGPS---R 218
OY 291 YAGSNLTALBAKTGLHQEVNLVWBRATOLQKNLTCEVWGPTSPKMLSLKLENKEAK 350
DB 219 FSGSG-----SGKDY---TLTITSLQAE-----238
OY 351 VSKREKPVWLNPEKGMQCLSDSGVLLSNNIKVLPWTSTPVPKPA-----PEPKS 403
DB 239 -----DPAITYC-----QQWSTPFTFGSGTELEIKVEKKS 269
OY 404 CDKHTTC-----PELLGSPVFLFPKPKDITLMSRTEPVCVVVDVSHEDPEVKFNYY 458
DB 270 SDKHTTCPCAPPELLGSPVFLFPKPKDITLMSRTEPVCVVVDVSHEDPEVKFNYY 329
OY 459 DGEVHNNAKTPREBOYNSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKTISKA 518
DB 330 DGEVHNNAKTPREBOYNSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKTISKA 389
OY 519 KGOFRPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDAVWESNGOENNKTTPPVLD 578
DB 390 KGOFRPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDAVWESNGOENNKTTPPVLD 449
OY 579 SDGSFFLYSKLTVDSKRMQGNVFSQVMHEALHNHYTKSLSPG 625
DB 450 SDGSFFLYSKLTVDSKRMQGNVFSQVMHEALHNHYTKSLSPG 496

```

## RESULT 63

US-10-363-427-24  
Sequence 24, Application US/10363427  
Publication No. US20030195338A1  
GENERAL INFORMATION:  
APPLICANT: MedexGen Inc.  
APPLICANT: CHUNG, Yong Hoon  
APPLICANT: HAN, Ji Woong  
APPLICANT: LEE, Hye Ja  
APPLICANT: CHOI, Eun Yong  
APPLICANT: KIM, Jin Mi  
APPLICANT: YIM, Soo Bin  
TITLE OF INVENTION: Concatameric Immunoadhesion  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/363,427  
CURRENT FILING DATE: 2003-02-28  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: Koparentin 1.71  
SEQ ID NO 24  
LENGTH: 502  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-363-427-24

Query Match 36.3%; Score 1239; DB 14; Length 502;

Best Local Similarity 47.9%; Pred. No. 2.4e-79;  
Matches 293; Conservative 40; Mismatches 111; Indels 168; Gaps 18;

```

OY 22 ATQGNKIVLGGKGDVELTCTASOKKSIQFHMKNKSNQIKIINGQSFITKPSKUNDPAD 81
DB 50 ATEVRVTVLRLRADSDQVEVCAATY-----MMGNELTF-----LDDISIC 87
OY 82 SRSILMDGNPPLIINKLKIEDSDTYICEVEDQKEVQLVFGTLTANSDTHLQGGSLTL 141
DB 88 TGTSSGNQVN--LITIGLRANDTGLYICKVE-----116
OY 142 TLESPPSSPSVOCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONOKKVEFKIDIV 201
DB 117 -LMPFP-----PYLIGNGTQGYVNDTE-----PCNDSNNHTAQ---PAVV 155
OY 202 LAFQKA-SIITYKKEGEVEFSFLATVEKLTSGELMOAERASSSKSWITFDLKNKE 260
DB 156 LASSRGIASFV-----CEYASPGKATEVAT---VLQADSDQVEVCAATY----- 198
OY 261 VSVKRVITQDPKLQMGKLLPLHLTPQALPQYAGSGNLTALBAKTGLHQEVNLVWBRAT 320
DB 199 -----MMGNELTF---LDDISICTGSSGN-----QVNLTIQGLR 229
OY 321 QLOKNL--TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLNPEAGMOCCLSDSGQVL 379
DB 230 AMDTGLYICKVELMVPPEYVIGI-----GNGTQYIVIDPE-----264
OY 380 LESNIKVLPTWSTVPKPA-APEPKSCDKHTTC-----PELLGSPVFLFPKPKDITLMS 433
DB 265 -----PCPDSAPKPSCKDHTCCPCAPPELLGSPVFLFPKPKDITLMS 309
OY 434 RTEPVCVVVDVSHEDPEVKFNMYVDGVEVHNNAKTPREBOYNSTYRVVSVLTVLHQMNL 493
DB 310 RTEPVCVVVDVSHEDPEVKFNMYVDGVEVHNNAKTPREBOYNSTYRVVSVLTVLHQMNL 369
OY 494 NGKEYKCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPY 553
DB 370 NGKEYKCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPY 429
OY 554 SDIAVWESNGOENNKTTPPVLDSDGSFFLYSKLTVDSKRMQGNVFSQVMHEALHN 613
DB 430 SDIAVWESNGOENNKTTPPVLDSDGSFFLYSKLTVDSKRMQGNVFSQVMHEALHN 489
OY 614 HYTKSLSLSPG 625
DB 490 HYTKSLSLSPG 501

```

## RESULT 64

US-10-404-724-8  
Sequence 8, Application US/10404724  
Publication No. US20030203447A1  
GENERAL INFORMATION:  
APPLICANT: Horwitz, Arnold H.  
TITLE OF INVENTION: Methods and Materials for Increasing Expression of Recombinant  
FILE REFERENCE: 13698US01  
CURRENT APPLICATION NUMBER: US/10/404,724  
CURRENT FILING DATE: 2003-03-31  
PRIOR APPLICATION NUMBER: US 60/368,530  
PRIOR FILING DATE: 2002-03-29  
NUMBER OF SEQ ID NOS: 79  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 8  
LENGTH: 465  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-404-724-8

Query Match 36.3%; Score 1238.5; DB 13; Length 465;  
Best Local Similarity 46.1%; Pred. No. 2.4e-79;  
Matches 296; Conservative 34; Mismatches 101; Indels 211; Gaps 17;



```

; CURRENT FILING DATE: 2003-02-03
; PRIOR APPLICATION NUMBER: US/09/733,764
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 09/590,656
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-357-653-2

```

```

Query Match 36.2% Score 1237; DB 14; Length 704;
Best Local Similarity 81.4% Pred. No. 5.3e-79;
Matches 240; Conservative 9; Mismatches 12; Indels 34; Gaps 5;

```

```

Qy 360 VLNPAGMOCCLSD-SCQVLLSNI--KVLPTWSTPVPCCAP----- 399
Db 414 ILPDSGVWCSVNTVAGVKEKFNISVKLPK-----PLNAPNVIDTGHNFVAINISSE 468
Qy 400 ----EPKSCDKTHTC-----PELLGSPVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDP 450
Db 469 PYFGSPKSCDKTHTCPCPAPPELLGSPVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDP 528
Qy 451 EVKFNMYVDGVEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKCKVSNKALPAP 510
Db 529 EVKFNMYVDGVEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKCKVSNKALPAP 588
Qy 511 IEKTSKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQGPENNY 570
Db 589 IEKTSKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQGPENNY 648
Qy 571 KTTPEVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVNHAEALHNHYTQKSLSLSPG 625
Db 649 KTTPEVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVNHAEALHNHYTQKSLSLSPG 703

```

RESULT 67

```

US-10-282-162-52
; Sequence 52, Application US/10282162
; Publication No. US20030143697A1
; GENERAL INFORMATION:
; APPLICANT: REGENERON PHARMACEUTICALS, INC.
; TITLE OF INVENTION: RECEPTOR BASED ANTAGONISTS, AND METHODS OF MAKING
; FILE REFERENCE: AND USING
; CURRENT APPLICATION NUMBER: US/10/282,162
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: 09/787,835
; PRIOR FILING DATE: 1999-09-22
; PRIOR APPLICATION NUMBER: PCT/US99/22045
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 52
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-282-162-52

```

```

Query Match 36.2% Score 1237; DB 14; Length 915;
Best Local Similarity 44.7% Pred. No. 7.5e-79;
Matches 293; Conservative 51; Mismatches 129; Indels 182; Gaps 18;
Qy 33 KGDVTELTCT-----ASOKSIOFHM-KNSNOIKILGNQSFLLTKGPSKLNDRADRR 84
Db 380 EGEPAALRCQVPPYWLMSVSRINLTMHKNDASRTVGG-----BEET 422
Qy 85 SLMDQGNPLIILKLIKIDSDTYICEVED-----QKEVQVLLVFGLTANSDTHL---LOGQ 137
Db 423 RMMADGALMLLPLAQ-EDSGTYVCTTNASGYCDKMSIELRVF---ENTDAFLPFIISYPQ 478

```

```

Qy 138 SLTLT-----LSPSPGS-----SPSVQCRSPRGKNIQGGKTLVSQLELQ 177
Db 479 ILTISTGSVLCPPLSETRTKTDVKIOWYDSDLLDNDNKFSLSVRGTLLVHVDVLE 538
Qy 178 DSGTWCTTV-----LONQKVEFKIDIVLAFOKASIVYKKEGVEVF 221
Db 539 DAGYRCVLTFAHGGQYNTIRSLIRIKKKEETIPVILISPLTISASLSR----- 591
Qy 222 SFPLATVEKLTGSGE-----LMMQAEASSSKSMITFDLKNKSVSVRYVQDPKLQMGK 276
Db 592 ---LTIPCKVLGTGTPLTMLMWTANDTHIESAV-----PGSRVTEGPRQEVSE 638
Qy 277 KLPLHLTPQALPOYAGSGNLTLLAEATGK-LHOEVNLVYMRATQLOKNLTCEVWCPTS 335
Db 639 NNENVIEVP-----LIFDPYTRDELHMDPRCVVHNTLSRQ----- 673
Qy 336 PKMLSLKLENKEAKVSKREKRVVNLNEAGMOCCLSDSGVLLSNIKVLPTWSTPVP 395
Db 674 ---TLRTTVKEAS----- 683
Qy 396 CPAPSPKSCDKTHTC-----PELLGSPVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDP 450
Db 684 ---STPSGDKTHTCPCPAPPELLGSPVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDP 739
Qy 451 EVKFNMYVDGVEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKCKVSNKALPAP 510
Db 740 EVKFNMYVDGVEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKCKVSNKALPAP 799
Qy 511 IEKTSKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQGPENNY 570
Db 800 IEKTSKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQGPENNY 859
Qy 571 KTTPEVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVNHAEALHNHYTQKSLSLSPG 625
Db 860 KTTPEVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVNHAEALHNHYTQKSLSLSPG 914

```

RESULT 68

```

US-10-679-620-62
; Sequence 62, Application US/10679620
; Publication No. US20040110930A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology
; APPLICANT: Reini, Stephen J
; TITLE OF INVENTION: MULTIMERIC PROTEIN ENGINEERING
; FILE REFERENCE: 34150-004A
; CURRENT APPLICATION NUMBER: US/10/679,620
; PRIOR FILING DATE: 2003-10-03
; PRIOR APPLICATION NUMBER: 60/415,940
; PRIOR FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 62
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: p9E10chimericv1-1, see Example 15
US-10-679-620-62

```

```

Query Match 36.2% Score 1236; DB 16; Length 715;
Best Local Similarity 46.4% Pred. No. 6.3e-79;
Matches 285; Conservative 39; Mismatches 104; Indels 186; Gaps 15;
Qy 30 LGGKDVTELTCTAS-----OKSIOFHKNSNOIKILGNQSFLLTKGPSKL 76
Db 269 LVKRGSLKLSKASGTFPSHYGMSWVRQTPDKRLN-----VATISRGTY-THYDSV 322
Qy 77 NDRADSRSLMDQGNPLIILKLIKIDSDTYICEVEDQKEVQVLLVFGLTANSDTHLQ 136
Db 323 KGRFTISRDN-DKNALYLOMNSLSKSEDTAMYC-----ARRSBRYYYGNTYYVSAMDYWG 376

```

```

QY 137 QSLTLTLESPSSSPVQCRSPKKNIOGKTLTSSVLOLSDSGTWTCTVLOKQKVERK 196
DB 377 QGASVTVSSASTKGPSPVFLPASPSSKSTSGG-TNALCGL----- 413
QY 197 IDIVLAFQKASSIVVKEGEGVEFSFPLAFIVEKLTGSGELMWQARASSSKMTTFDL 256
DB 414 -----VKDYFPEPVTVS-----MNSGALTSG----- 434
QY 257 KKEKSVKRVTPDKLQMGKLPFLHLTLPOALPOYVAGSGLTLALPAKTKLHOEVNLVY 316
DB 435 -----VH-TFPAVL-QSSGLTSSSVTVVPSSTLGTQYI-- 467
QY 317 MRATQLOKQNLTCGWGPTSPKMLSLKLENKEKAVSKREKPVVNLPEAGMQLCLSDSG 376
DB 468 -----CNV-----NHKPSNTKVDKRV----- 483
QY 377 QVLLSNIKVLPTWSTPVPCPAPEPKSCDKTHTC-----PELLGSPVFLFPKPKDTLM 431
DB 484 -----EPKSCDKTHTCPCPAPELLGSPVFLFPKPKDTLM 520
QY 432 ISRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYRVSVLTVLHOD 491
DB 521 ISRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYRVSVLTVLHOD 580
QY 492 WLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGF 551
DB 581 WLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGF 640
QY 552 YPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGNVFSCSVMHEAL 611
DB 641 YPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGNVFSCSVMHEAL 700
QY 612 HNNHYTKSLSLSPG 625
DB 701 HNNHYTKSLSLSPG 714

RESULT 69
US-10-683-255-6
; Sequence 6, Application US/10683255
; Publication No. US20040063910A1
; GENERAL INFORMATION:
; APPLICANT: Kavanaugh, William M.
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR
; FILE REFERENCE: PP01474.101
; CURRENT APPLICATION NUMBER: US/10/683,255
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 09/499,846
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/119,002
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-683-255-6

Query Match 36.2%; Score 1235.5; DB 12; Length 497;
Best Local Similarity 66.2%; Pred. No. 4.2e-79;
Matches 255; Conservative 18; Mismatches 54; Indels 58; Gaps 8;

QY 293 GSGNLTALAEAKTKLHQ---EVLNVVMRATQLOKNT--TCEVWGPTSPKMLSLKLE-- 345
DB 118 GSIHNTYQLDVVERSHRPILOGLPANKTVAGSNVPEMCKKYSDDQPHIQMLKHIEVN 177
QY 346 -----NKEAKVSKREKPVVNLN-----PEAGMQLCLSDSGQVLLSNIK 385
DB 178 GSKIGPDLNLPYVQILTKTAGVNTTDEKMEVHLRLNVSFEDAGEYTCLAGNSIGLSHHS--- 234

```

```

QY 386 VLPTWST-----VPCPAP-----EPKSCDKTHTC-----PELLGSPVFL 420
DB 235 ---AMLTVLEALEERPAWMTSPLYLEGSSPGLQEPKSCDKTHTCPCPAPELLGSPVFL 291
QY 421 LFPKPKDTLMSRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYR 480
DB 292 LFPKPKDTLMSRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYR 351
QY 481 VVSVTLVHODWLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 540
DB 352 VVSVTLVHODWLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 411
QY 541 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGN 600
DB 412 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGN 471
QY 601 VFSCSVMEHALHNNHYTKSLSLSPG 625
DB 472 VFSCSVMEHALHNNHYTKSLSLSPG 496

RESULT 70
US-10-683-255-4
; Sequence 4, Application US/10683255
; Publication No. US20040063910A1
; GENERAL INFORMATION:
; APPLICANT: Kavanaugh, William M.
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR
; FILE REFERENCE: PP01474.101
; CURRENT APPLICATION NUMBER: US/10/683,255
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 09/499,846
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/119,002
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-683-255-4

Query Match 36.2%; Score 1235.5; DB 12; Length 525;
Best Local Similarity 66.2%; Pred. No. 4.5e-79;
Matches 255; Conservative 18; Mismatches 54; Indels 58; Gaps 8;

QY 293 GSGNLTALAEAKTKLHQ---EVLNVVMRATQLOKNT--TCEVWGPTSPKMLSLKLE-- 345
DB 146 GSIHNTYQLDVVERSHRPILOGLPANKTVAGSNVPEMCKKYSDDQPHIQMLKHIEVN 205
QY 346 -----NKEAKVSKREKPVVNLN-----PEAGMQLCLSDSGQVLLSNIK 385
DB 206 GSKIGPDLNLPYVQILTKTAGVNTTDEKMEVHLRLNVSFEDAGEYTCLAGNSIGLSHHS--- 262
QY 386 VLPTWST-----VPCPAP-----EPKSCDKTHTC-----PELLGSPVFL 420
DB 263 ---AMLTVLEALEERPAWMTSPLYLEGSSPGLQEPKSCDKTHTCPCPAPELLGSPVFL 319
QY 421 LFPKPKDTLMSRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYR 480
DB 320 LFPKPKDTLMSRTPEVTCVVDVSHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYR 379
QY 481 VVSVTLVHODWLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 540
DB 380 VVSVTLVHODWLNKGEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 439
QY 541 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGN 600
DB 440 QVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQGN 499

```



Qy 601 VFSCSVMEALHNHYTOKSLSPG 625  
Db 500 VFSCSVMEALHNHYTOKSLSPG 524

RESULT 71  
US-10-207-655-397  
; Sequence 397, Application US/10207655  
; Publication No. US20030118592A1  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Hayden-Ledbetter, Martha S.  
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
; FILE REFERENCE: 390069.401C1  
; CURRENT APPLICATION NUMBER: US/10/207,655  
; PRIOR FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 426  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 397  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion polypeptide  
US-10-207-655-397

Query Match 36.1%; Score 1234; DB 14; Length 500;  
Best Local Similarity 47.6%; Pred. No. 5.4e-79;  
Matches 294; Conservative 37; Mismatches 128; Indels 158; Gaps 18;  
Qy 23 TQGNKVVVGKGGDPTVELTCTASQKSIQFMKNSNQIKLNGSGFLTKGSKLNDRAD 82  
Db 27 SQSRAILSASFGCEKVTMTCRASSVS-YMHYQOKP-----GSSRPWTYASNLASGPA 81  
Qy 83 RRSIMDQG-NPPLIIKNLIKEDSDTYICEVEDQKEVQLLVFGLTANSSTHLQGSITL 141  
Db 82 RPSGSGSGTSLTISRVEAEDAATYYCO-----QMSFNPTF---GAGTKL 125  
Qy 142 TLSEPRGSPVQCRSPRGKNIQGGKITSVQLBELQDGTCTYVLQKQKVEPRIDIV 201  
Db 126 ELKGGGGG-----GGSGGGGS---SQAYLQSGA-----ELV- 156  
Qy 202 LAFQASSIVYKKEGEQVFSFPLAFVTEKLTG-----SGELMQAERASSSKMTFPLK 257  
Db 157 -----REGASVAKSC-----KASGTFITSYNMHWAKQTRQGLEMIG----- 193  
Qy 258 NKEVSVKRVTDPKLQMGKLLPLHLTPQALPOYAGSGNLTALAEAKTGKLEHVNLYVM 317  
Db 194 -----AIVPNGDTSYNQFK-GK----- 211  
Qy 318 RATOLQKRLTCEWGPSPKMLSLKLEKAKVSKREKPVWVLPKAGMQLSDSGQ 377  
Db 212 -----ATLTVDKSSSTAYMQLSLTSE-----DSAVVFC-----AR 242  
Qy 378 VLLESN-----IKVLPTASTPVPCCAPRPEKCDKTHT-----CPELLGSPVFLPPPKPD 428  
Db 243 VVYISNSWYRPDWGCTGTTVVSSDQPKSCDKHTHTSPCCAPRPLGGSPVFLPPPKPD 302  
Qy 429 TLMISRPPEVTCVVVDVSHEDPEVKFMNYVDGVEVHNAKTRPREQVNSTYRVVSVLTVL 488  
Db 303 TLMISRPPEVTCVVVDVSHEDPEVKFMNYVDGVEVHNAKTRPREQVNSTYRVVSVLTVL 362  
Qy 489 HQDWLNGEKYCKYKSNKALPAPIETKISKAGQPREPQVYTLPPSRDELTKNQVSLTCLV 548  
Db 363 HQDWLNGEKYCKYKSNKALPAPIETKISKAGQPREPQVYTLPPSRDELTKNQVSLTCLV 422  
Qy 549 KGFPSPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDSKRWQGNVSCSYMH 608  
Db 423 KGFPSPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDSKRWQGNVSCSYMH 482  
Qy 609 EALHNHYTOKSLSPG 625  
Db 483 EALHNHYTOKSLSPG 499

RESULT 72  
US-10-452-646-9  
; Sequence 9, Application US/10452646  
; Publication No. US20040018593A1  
; GENERAL INFORMATION:  
; APPLICANT: Cartoon, Jill M.  
; APPLICANT: Steadnet, Kimberly C.  
; APPLICANT: Scallion, Bernard J.  
; APPLICANT: Jili, Giles-Komar  
; TITLE OF INVENTION: ANTI-RELP FUSION ANTIBODIES, COMPOSITIONS, METHODS AND USES  
; FILE REFERENCE: CEN0296 NP  
; CURRENT APPLICATION NUMBER: US/10/452,646  
; PRIOR FILING DATE: 2003-06-02  
; PRIOR APPLICATION NUMBER: US 60/385,305  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 9  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-452-646-9

Query Match 36.1%; Score 1233; DB 15; Length 367;  
Best Local Similarity 74.7%; Pred. No. 4.2e-79;  
Matches 251; Conservative 17; Mismatches 30; Indels 38; Gaps 10;  
Qy 326 LTCVWGPSTSP-KLMLSLKLEKAKV---SKREKPVWV-LNPE-AGMQL----- 371  
Db 33 LECQSYNGAHLASLILKEASTIAEYISGYRQSPFIWIGHDPQKQWQWIDGAMWLY 92  
Qy 372 LSDSGVLT-----LESNIKVLPTWSTPVPCC-----APRPSCKTHTC----- 410  
Db 93 RSMGSKMGNGKHAEMSSNNFL-TWSSN-ECNKROHFLCKYRPEPKSCDKHTCTCPCCP 150  
Qy 411 -PELLGSPVFLPPPKPDITLMSRTPREVTCCVVVDVSHEDPEVKFMNYVDGVEVHNAKTK 469  
Db 151 APELLGGSVFLPPPKPDITLMSRTPREVTCCVVVDVSHEDPEVKFMNYVDGVEVHNAKTK 210  
Qy 470 PREQVNSTYRVVSVLTVLHQDWLNGEKYCKYKSNKALPAPIETKISKAGQPREPQVYTL 529  
Db 211 PREQVNSTYRVVSVLTVLHQDWLNGEKYCKYKSNKALPAPIETKISKAGQPREPQVYTL 270  
Qy 530 LPPSRDELTKNQVSLTCLVKGFPSPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKL 589  
Db 271 LPPSRDELTKNQVSLTCLVKGFPSPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKL 330  
Qy 590 TVDSKRWQGNVSCSYMHLEALHNHYTOKSLSPG 625  
Db 331 TVDSKRWQGNVSCSYMHLEALHNHYTOKSLSPG 366

RESULT 73  
US-10-207-655-396  
; Sequence 396, Application US/10207655  
; Publication No. US20030118592A1  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Hayden-Ledbetter, Martha S.  
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
; FILE REFERENCE: 390069.401C1  
; CURRENT APPLICATION NUMBER: US/10/207,655  
; PRIOR FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 426  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 396  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion polypeptide

US-10-207-655-396

Query Match 36.1%; Score 1233; DB 14; Length 500;  
 Best Local Similarity 47.6%; Pred. No. 6,4e-79;  
 Matches 294; Conservative 37; Mismatches 128; Indels 158; Gaps 18;

23 TOGKVVLLGKGGTVELTCTASQKSIQFHWKNSNQIKLGNQGSFLTQKPSKLNDRADS 82  
 27 SOSPAIILASPEKGVYTCRASSSVS-YMHMYQKP---GSSPKWYIAPSNLASGVPA 81  
 83 RSLMDQG-NPFLIIKULKIEDSDTYICEVEDQKEVQLVFGILTANSDTHLQGSFLT 141  
 82 RFGSGSGTSYSLTISVEAEDATYYCQ-----QMSNPPTF---GAGTGL 125  
 142 TLESPPGSSPSVOCSPERKNIQGGKTLISVQLFLDSDGTWCTVVLQNKQKFEKIDIV 201  
 126 ELKDGGGSG-----GGSGGGGSG---SQAYLQGSQA-----ELV- 156  
 202 LAFQKASSIYKKEGQVEFSPLAFVTEKLTG---SGELMQAERASSSKSWITFDLK 257  
 157 -----RFGASVKNMSC-----KASGTYTTSYMHVYKQTPROGLEWIG--- 193  
 258 NKEVSVKRVTDPRKLOMGKLLPLHLTLPCALPOYAGSGNLTALAEKTKLHOEVNLVYM 317  
 194 -----AIPGNGDTSYNQKFK-GK----- 211  
 318 RATQLOKNTLCEVWGPSTPKLMLSLKLENKAKVSKREKFWVLNPEAGMOCLLSDSQ 377  
 212 -----ATLTVDKSSSTAYWQLSSLTSE-----DSAVYFC-----AR 242  
 378 VLLESN-----IKVLPMTSTVPVPCAPPEKSCDKTHTC-----PELLGSPVFLPPPKPD 428  
 243 VVYSSSYMYFPDWGVTGTTTVSSDDEPKSSDKTHTCPCPAPBELGGSPVFLPPPKPD 302  
 429 TLMISRTPEVTCVVDVSHEDPEVKFNMVYDGVENVNAKTPREBOYNSTYRVVSVLTVL 488  
 303 TLMISRTPEVTCVVDVSHEDPEVKFNMVYDGVENVNAKTPREBOYNSTYRVVSVLTVL 362  
 489 HQMLNGKEYKCYKSNKALPAPIEKTSKAKGQPREQVYTLPPSRDELTKNOVSLTCLV 548  
 363 HQMLNGKEYKCYKSNKALPAPIEKTSKAKGQPREQVYTLPPSRDELTKNOVSLTCLV 422  
 549 KGYPSDIAVEMESNGQPENNYKTPPVLDSDSFFLYSKLTVDKSRMOQGNVFCGSVMH 608  
 423 KGYPSDIAVEMESNGQPENNYKTPPVLDSDSFFLYSKLTVDKSRMOQGNVFCGSVMH 482  
 QY 609 EALHNYTKSLSPG 625  
 DB 483 EALHNYTKSLSPG 499

RESULT 74  
 US-10-363-427-20  
 ; Sequence 20, Application US/10363427  
 ; Publication No. US20030195338A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MedexGen Inc.  
 ; APPLICANT: CHONG, Yong Hoon  
 ; APPLICANT: HAN, Ji Woong  
 ; APPLICANT: LEE, Hye Ja  
 ; APPLICANT: CHOI, Eun Yong  
 ; APPLICANT: KIM, Jin Mi  
 ; APPLICANT: YIM, Soo Bin  
 ; TITLE OF INVENTION: Concatameric Immunoadhesion  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/363,427  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: Kopatentin 1.71  
 ; SEQ ID NO 20  
 ; LENGTH: 502  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

US-10-363-427-20

Query Match 36.1%; Score 1233; DB 14; Length 502;  
 Best Local Similarity 68.5%; Pred. No. 6,4e-79;  
 Matches 250; Conservative 12; Mismatches 43; Indels 60; Gaps 7;

290 QVAGSGNLTALAEATGKLGHOEVNLVYMRATQLOKNTLCTC-----EYWGPT 334  
 168 EYASFGKAT-----EYRTVTLKQADSVTEVCATATYMMGNELTFLDSDICTGT 215  
 335 SPKMLSLKENKA-----KVSKEKPVVNLPEAGMOCLLSDSGOVLLESNIKY 386  
 216 SSGNQVNLITQGLRAMDITGLYICKVLMVPPYYIGIGNG-----TQIYV 260  
 387 LPTWSTVPCP-APEPKSCDKTHTC-----PELLGSPVFLPPPKDITLMSRTPEVTC 440  
 261 I-----DPEPCPSASPEKSCDKTHTCPCPAPBELGGPSVFLPPPKDITLMSRTPEVTC 316  
 441 VVVDVSHEDPEVKFNMVYDGVENVNAKTPREBOYNSTYRVVSVLTVLHQDMLNGKEYKC 500  
 317 VVVDVSHEDPEVKFNMVYDGVENVNAKTPREBOYNSTYRVVSVLTVLHQDMLNGKEYKC 376  
 501 KVSNNKALPAPIEKTSKAKGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEM 560  
 377 KVSNNKALPAPIEKTSKAKGQPREQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEM 436  
 561 ESNQGPENNYKTPPVLDSDSFFLYSKLTVDKSRMOQGNVSCSYMHKALHNYTKSL 620  
 437 ESNQGPENNYKTPPVLDSDSFFLYSKLTVDKSRMOQGNVSCSYMHKALHNYTKSL 496  
 QY 621 SLSPG 625  
 DB 497 SLSPG 501

RESULT 75  
 US-10-207-655-346  
 ; Sequence 346, Application US/10207655  
 ; Publication No. US20030118592A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ledbetter, Jeffrey A.  
 ; APPLICANT: Hayden-Ledbetter, Martha S.  
 ; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
 ; FILE REFERENCE: 390069, 401C1  
 ; CURRENT APPLICATION NUMBER: US/10/207,655  
 ; CURRENT FILING DATE: 2002-07-25  
 ; NUMBER OF SEQ ID NOS: 426  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 346  
 ; LENGTH: 543  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: fusion polypeptide  
 US-10-207-655-346

Query Match 36.1%; Score 1233; DB 14; Length 543;  
 Best Local Similarity 46.0%; Pred. No. 7,1e-79;  
 Matches 296; Conservative 43; Mismatches 115; Indels 190; Gaps 19;

1 MNRGVFRRLLVLAQLALPAATQGNKVVLGKGGTVELTCTASQKSIQFHWKNSNQIK 60  
 19 MSKGVN-----IVL-----TQSPITIASBGEVNTTTCRASSSVSYMYWYQKS--- 62  
 61 ILGNQGSFLTQKPSKLNDRADSRSLMDQ-NPFLIIKULKIEDSDTYICEVEDQKEVQ 119  
 63 --GASPKMLTYDPSKLAGVPMRFGSGSGTSYSLAINMTEFTDATYYCQ----- 111  
 120 LNVFGILTANSDTHLQGSFLTLESPPGSSPSVOCSPERKNIQGGKTLISVQLFLDSD 179  
 112 -----QMSSTPLTF-----GSGTGLDIKRGGGGSGGGGSGQVOLDKEA 152  
 QY 180 G-----TWCTVVLQNKQKFEKIDIVLAFQKASSIYKKEGQVEFSPLAFV 229

```

Db      153 GPGIIVPTQTSLTCTV-----SGFS-----173
Qy      230 EKLIGSELWQBARASSKSWITFDLKNKEVSKRVYTOPDKLQMGKPLPLHLTPQALP 289
Db      174 --LTSVGVHW-----IRQP---GKLEW---MGII 196
Qy      290 QYASSGNLTLLAEKTKGLHGVNLVVMRATOLQKNTLCEVWGPTSPKLMSTLKENKA 349
Db      197 YVDCGTQYNSAIKSR-----LSTSRDS-----KSYFLKINSIQ- 231
Qy      350 KVSREKRVWVNLNPEAGWQCL---LSDSGVLLSNIKVLPTWSTVPCCAPEPKSCDK 406
Db      232 -----TDDTAMYCARLHPDYWGQ-----GVWTVVSSDL-----EPKSSDK 267
Qy      407 THT-----CPELLGSPVFLFPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVDGV 461
Db      268 THSPSPAPPELLGSSVFLFPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVDGV 327
Qy      462 EVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYCKVSNKALPAPIEKTISKAKGQ 521
Db      328 EVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYCKVSNKALPAPIEKTISKAKGQ 387
Qy      522 PREQVYVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTPEVLDSDG 581
Db      388 PREQVYVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTPEVLDSDG 447
Qy      582 SFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 625
Db      448 SFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 491

```

## RESULT 76

```

US-10-107-991B-3
; Sequence 3, Application US/10107991B
; Publication No. US20040058445A1
; GENERAL INFORMATION:
; APPLICANT: LEDBETTER, JEFFREY
; APPLICANT: HAYDEN-LEDBETTER, MARTHA
; APPLICANT: HELSTROM, INGBERD
; APPLICANT: HELSTROM, KARL ERIK
; TITLE OF INVENTION: ACTIVATION OF TUMOR-REACTIVE LYMPHOCYTES VIA ANTIBODIES
; FILE REFERENCE: 034474.0004
; CURRENT APPLICATION NUMBER: US/10/107, 991B
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 60/286, 585
; PRIOR FILING DATE: 2001-04-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 2.1
; SEQ ID NO 3
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mouse-Human Fusion Protein
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(23)
; OTHER INFORMATION: L6 V kappa signal peptide
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (24)..(133)
; OTHER INFORMATION: G19-4 mouse anti-human CD3 light chain variable domain
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (134)..(148)
; OTHER INFORMATION: (Gly4Ser)3 linker peptide
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (149)..(270)
; OTHER INFORMATION: G19-4 mouse anti-human VH domain
; FEATURE:

```

```

; NAME/KEY: MISC FEATURE
; LOCATION: (271)..(504)
; OTHER INFORMATION: human IgG1 Fc domain (hinge, CH2, CH3)
; FEATURE:
; NAME/KEY: TRANSMEM
; LOCATION: (505)..(555)
; OTHER INFORMATION: human CD80 transmembrane domain and cytoplasmic tail
US-10-107-991B-3

```

```

Query Match      36.1%; Score 1233; DB 12; Length 555;
Best Local Similarity 46.3%; Pred. No. 7,3e-79;
Matches 299; Conservative 42; Mismatches 123; Indels 182; Gaps 20;

```

```

Qy      1  MNRGVPRHLVLVQLALPLAATQGNKVVGLKGGDTVELCTASQKSIQFHW---KNSN 57
Db      19  MSRGVDIQ-----MTQTSLSASLQDRRTISCRASQDRIANLWYQOQFPDG 65
Qy      58  QIKILNQSGSFLTKGPSKLNDRADSRSLWDG--NPLIINKLKIEDSDTYICEVDQKE 116
Db      66  TVKLL-----LYYT---SRHSVPSRFSGSGSGTDYSLTIANLPEDIATYFCQ----- 112
Qy      117  EVQLLVGLTANSPTHLLOGQSLLTLESPPGSSPSYQCSPRGKNIQGGKT---LSVSG 173
Db      113  -----QNTLPWTF---GGGTKLVTKRELGGSGSGSGSGSGSIDE 151
Qy      174  LEIQQSGTWTCTVLQNGKVFKIDIVLAFQKASSIVYKKEGQVFFPLAFTVEKLT 233
Db      152  VQLQSGP-----ELV---KQASNSCASG---YSF-TGYIV--- 183
Qy      234  GSGELMWQBARASSKSWITFDLKNKEVSKRVYTOPDKLQMGKPLPLHLTPQALPYAG 293
Db      184  -----WLQSHGKNLEIGLINPYKGLT---TYNQKFK----- 213
Qy      294  SGNLTLLAEKTKGLHGVNLVVMRATOLQKNTLCEVWGPTSPKLMSTLKENKAAYSK 353
Db      214  -GKATLTVDKSSSTAYWE-----LSTSEDSAAVYYCA 245
Qy      354  REK-----PYVNLNPEAGWQCLLSDSGVLLSNIKVLPTWSTVPCCAPEPKSC 404
Db      246  RSGYGDSDMYFDW---GAGTTVTYSSDL-----EPKSS 277
Qy      405  DKHT-----CPELLGSPVFLFPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVD 459
Db      278  DKHTSPSPAPPELLGSSVFLFPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVD 337
Qy      460  GVEVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYCKVSNKALPAPIEKTISKAK 519
Db      338  GVEVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYCKVSNKALPAPIEKTISKAK 397
Qy      520  GQPREPVYVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTPEVLD 579
Db      398  GQPREPVYVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTPEVLD 457
Qy      580  DGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 625
Db      458  DGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 503

```

## RESULT 77

```

US-10-363-427-16
; Sequence 16, Application US/10363427
; Publication No. US20030195338A1
; GENERAL INFORMATION:
; APPLICANT: MedexGen Inc.
; APPLICANT: CHUNG, Yong Hoon
; APPLICANT: HAN, Ji Woong
; APPLICANT: LEE, Hye Ja
; APPLICANT: CHOI, Eun Yong
; APPLICANT: KIM, Jin Mi
; APPLICANT: YIM, Soo Bin
; TITLE OF INVENTION: Concatametric Immunoadhesion
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/363, 427

```

```

; CURRENT FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 16
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-363-427-16

Query Match      36.1%; Score 1232; DB 14; Length 377;
Best Local Similarity 68.8%; Pred. No. 5.1e-79;
Matches 251; Conservative 12; Mismatches 42; Indels 60; Gaps 7;

QY 290 QYAGSNTLALAEAKGKGLHGVNLYVMRATQLOKMLTC-----EVMGPT 334
DB 43 EVASPKAT-----EVRVTVLRQADSQVTEVCATYMGNELTFLDLSICTGT 90
QY 335 SPKMLSLKLENKA-----KVSREKPVVWVLPBAGMQLLSDSGVLLSEINIKY 386
DB 91 SSGNQVNLITIGLRANDTGLYICKVELMPPYYLIGNG-----TQIVY 135
QY 387 LPTWSTPVPBP-APEPKSCDKHTTC-----PELLGSPVFLPFPKPKDTLMTSRTEVTC 440
DB 136 I-----DPEPCPDABEPKSCDKHTTCPCCPAPPELLGSPVFLPFPKPKDTLMTSRTEVTC 191
QY 441 VVVVDVSHEDPEVKFNNYVDGVEVHNAKTKRBEQYNSTRYVSVLTVLHODMVLNGEKYK 500
DB 192 VVVVDVSHEDPEVKFNNYVDGVEVHNAKTKRBEQYNSTRYVSVLTVLHODMVLNGEKYK 251
QY 501 KVSNNKLPAPLEKTIKSKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW 560
DB 252 KVSNNKLPAPLEKTIKSKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW 311
QY 561 ESNQGENNYKTPPVLDSDGSFELYSKLTVDKSRMQGNVFSQVMEALHNHYTQKSIL 620
DB 312 ESNQGENNYKTPPVLDSDGSFELYSKLTVDKSRMQGNVFSQVMEALHNHYTQKSIL 371
QY 621 SLSPG 625
DB 372 SLSPG 376

RESULT 78
US-10-334-235-38
; Sequence 38; Application US/10334235
; Publication No. US20040131591A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Ltd.
; APPLICANT: Kingman, Alan
; APPLICANT: Bebbington, Christopher
; APPLICANT: Carroll, Miles
; APPLICANT: Ellard, Fiona
; APPLICANT: Kingman, Susan
; APPLICANT: Myers, Kevin
; APPLICANT: Lamikandra, Abigail
; TITLE OF INVENTION: VECTOR SYSTEM
; FILE REFERENCE: 53268200920
; CURRENT APPLICATION NUMBER: US/10/334,235
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 10/060,585
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: PCT/GB00/04317
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 09/445,375
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptide of 575Aab1

```

```

US-10-334-235-38

Query Match      36.1%; Score 1231; DB 16; Length 600;
Best Local Similarity 44.7%; Pred. No. 1.1e-78;
Matches 302; Conservative 49; Mismatches 131; Indels 194; Gaps 21;

QY 30 LKKKGDVTELTCTAS--QKKSIOFHMKNKSNQKIIG-----NQGSLTGPSTLNDRA-- 80
DB 33 LVKPGASVKISCKASGYSTGYMHWVKSHKSKSLFEMIRIPNNQVTLTYNQKFKDKAIL 92
QY 81 ---DSRRSLMDQGNPPLIINKLIKEDSPDYICE-----VEDQKEVQLLVFG----- 124
DB 93 TVDKSSTIAY-----MELRSLTSEDSAYVCARSMITNNYMDWVGQVTSVTVSSGGGG 146
QY 125 -----LTRANSITLL--QGOSLTLT-----LSBPQSSPS--VQ 154
DB 147 SGGGGTGGGSSIVWTFPTFLVAGDVITTCASQVSNVAVVQKPGQSPFLTIS 206
QY 155 CSRPR-----QKNIQGGKTLVSQLEQDSGTMTCTVLONO-----KKVEFKID 198
DB 207 YTSRIRAGVDPDFISGCTGDTFTITLQAEIDLAVFQODYNSPPTGGGKLEIK-- 264
QY 199 IVVLAFOKASSIVYKKEGQVEFSEPLAFTVEKLTSGGELMWQAERASSSKSWITFDLKN 258
DB 265 -----RAST-----KGPSV--FPL-----APSSKS----- 282
QY 259 KEVSVKRVYQDPKLOMGKKLPLHLPLPOLPYAG--SGNLTALAEATGKLGHOEVNLY 315
DB 283 -----TSGGTALGCLVTKDYPPEPVTVSNWNGALTSGVHTFPVAVLQSSGLYSLSVY 334
QY 316 VNRATQL-QKNLTCEVWGTPSPKMLSLKENKEKAVSREKPVWVLPBAGMQLLSD 374
DB 335 TVPSSSLGTQYICNV-----NHSKSNITVDKXV----- 363
QY 375 SGQVLLSESNIKVLPWSTPVPBPAPPEPKSCDKHTTC-----PELLGSPVFLPFPKPKDT 429
DB 364 -----EPKSCDKHTTCPCCPAPPELLGSPVFLPFPKPKDT 398
QY 430 LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRBEQYNSTRYVSVLTVLH 489
DB 399 LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRBEQYNSTRYVSVLTVLH 458
QY 490 QDMINGEKYKCVSNKALPAPLEKTIKSKAKQPREPOVYTLPPSRDELTKNQVSLTCLVK 549
DB 459 QDMINGEKYKCVSNKALPAPLEKTIKSKAKQPREPOVYTLPPSRDELTKNQVSLTCLVK 518
QY 550 GFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFELYSKLTVDKSRMQGNVFSQVME 609
DB 519 GFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFELYSKLTVDKSRMQGNVFSQVME 578
QY 610 ALHNHYTQKSLSLSPG 625
DB 579 ALHNHYTQKSLSLSPG 594

RESULT 79
US-10-435-299-7
; Sequence 7; Application US/10435299
; Publication No. US20040052783A1
; GENERAL INFORMATION:
; APPLICANT: Weinert, George
; APPLICANT: Gingrich, Roger
; APPLICANT: Link, Brian
; APPLICANT: Tso, J. Yun
; TITLE OF INVENTION: HUMANIZED ANTIBODIES AGAINST CD3
; FILE REFERENCE: 05882-0176-CNUS04
; CURRENT APPLICATION NUMBER: US/10/435,299
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 09/618,380
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 08/397,411
; PRIOR FILING DATE: 1995-03-01
; PRIOR APPLICATION NUMBER: US 07/859,583

```

```

; PRIOR FILING DATE: 1992-03-27
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Complete heavy chain of Humanized 1D10 Ab
US-10-435-299-7

```

```

Query Match      36.0%; Score 1230.5; DB 12; Length 446;
Best Local Similarity 46.7%; Pred. No. 8.2e-79;
Matches 284; Conservative 36; Mismatches 103; Indels 185; Gaps 14;

```

```

QY 30 LGKGDVVELTCTSAQKSIQF--HMKNSNQIKILGNQSFYTGKPSKLNDRADSRSL- 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 11 LKPSSETLSTCTGSGFLTNVGVWVRQSPKGLGWVGMQSGSTEYNAAFISRLTIS 70
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 87 --WPGNPELIIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQ--GQSLLT 142
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 71 KDTSKNQVSLKLNSTLTADTAATYC-----AKNDRYAMDYWGQGLVLT 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 143 LESPFGSSPSVQCSRPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNKVKEFKIDIVVL 202
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 114 VSSASTKGPSVFLPAPSSKTSQ--TALGCL----- 144
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 203 AFQKASSIVYKKEGQVEFPLAFTVEKLTGSGELMWQAEERASSSKSWITFDLKNKVS 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 145 --VDYFPEPVTVS-----WNSGALTSG----- 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 263 VKRVTQDPKLGKKLPLHLTLPLQALPOYAGSGNLTLLAEKTKLHQEVNLVYMRATQL 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 166 -----VH-TFPAVL--QSSGLYSLSSVTVPSLSLGTQTYI----- 198
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 323 QKNLTCEVWGPSTPRLMLSLKLENKAKVSKREKPVWVLNPEAGMQCLSDSGQVLLS 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 199 -----CNV-----NHKPSNTKVDKKV----- 214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 383 NIKVLPWTSTVPCPAPPSKCDKTHTC-----PELLGSPVFLPPPKDITLMSRTE 437
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 215 -----EPKSCDKHTHPCPAPPELLGSPVFLPPPKDITLMSRTE 257
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 438 VTCVVDVSHDEPVEKFMVYDGVGVHNAKTPREEOVNSTYRVVSVTLVHQMNGKE 497
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 258 VTCVVDVSHDEPVEKFMVYDGVGVHNAKTPREEOVNSTYRVVSVTLVHQMNGKE 317
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 498 YKCAVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIA 557
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 318 YKCAVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIA 377
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 558 VEMESNGQPENNYKTPPVLDSGFFLYSKLTVDKSRMQGNVFGSCVMHEALAHNYQT 617
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 378 VEMESNGQPENNYKTPPVLDSGFFLYSKLTVDKSRMQGNVFGSCVMHEALAHNYQT 437
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 618 KSLSLSPG 625
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 438 KSLSLSPG 445

```

```

RESULT 80
US-10-207-655-270
; Sequence 270, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Ledbetter, Jeffrey A.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0

```

```

; SEQ ID NO 270
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion polypeptide
US-10-207-655-270

```

```

Query Match      36.0%; Score 1230.5; DB 14; Length 550;
Best Local Similarity 45.1%; Pred. No. 1.1e-78;
Matches 299; Conservative 35; Mismatches 108; Indels 221; Gaps 19;

```

```

QY 1 MNRGVPRHLLVQLALLPAAQGNKVLGKKDVELTCTASQ--KKS-I-QFMKNSN 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 19 MSRGVDIQ-----MTQSPSSLSASLGKTTIKTSQDIIKTIQWQHPRK 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 58 QIKILGNQSFYTGK--PSKLNDRADSRSLMDQNPFLIIKNIKIEDSDTYICEVEDQKE 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 66 GPRLLIYVSTLQGISRFGSGSGR-----DYSLSIRNLEPEDIAATYQ----- 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 117 EVQLVFGTANSDTHLQGSLLTLESPFGSSPSVQCSRPRGKNIQGGKTLVSQLEL 176
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 113 -----QYDNLPLTF-----GSGTKLEIRGGGSGGSGGSDVQL 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 177 QDSG-----TWTCVTLQNKVKEFKIDIVLAFQKASSIVYKKEGQVEFSPPLA 226
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 151 QDSGFLVKSQSLSLTCV----- 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 227 FVEKLTG-----SGELM--WQAEERASSSKSW--ITFDLKN-----KEVSVKRVTDPK 271
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 171 -----TGYSITSGFYNNMIRQFPGNKLEWNGHISHDGRNNVPSLINRISITRDT--K 222
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 272 LQNGKKLPLHLTLPLQALPOYAGSGNLTLLAEKTKLHQEVNLVYMRATQLQKNLT 327
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 223 NQFELKLSVTTEDTATYFCARHNGSGAM----- 252
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 328 CEVWGPSTPRLMLSLKLENKAKVSKREKPVWVLNPEAGMQCLSDSGQVLLSNIKVL 387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 253 -DYWG----- 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 388 PTWSTVPCPAPPSKCDKTHTC-----CPPELLGSPVFLPPPKDITLMSRTEVTCV 442
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 268 -----EPKSCDKHTHSPAPPELLGSSVFLPPPKDITLMSRTEVTCV 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 443 VDVSHDEPVEKFMVYDGVGVHNAKTPREEOVNSTYRVVSVTLVHQMNGKEKYCKV 502
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 316 VDVSHDEPVEKFMVYDGVGVHNAKTPREEOVNSTYRVVSVTLVHQMNGKEKYCKV 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 503 SNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEMES 562
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 376 SNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEMES 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 563 NGQPENNYKTPPVLDSGFFLYSKLTVDKSRMQGNVFGSCVMHEALAHNYTQKSL 622
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 436 NGQPENNYKTPPVLDSGFFLYSKLTVDKSRMQGNVFGSCVMHEALAHNYTQKSL 495
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 623 SPG 625
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 496 SPG 498

```

```

RESULT 81
US-10-418-836-38
; Sequence 38, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836

```

```

; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-38

```

Query Match 36.0%; Score 1230.5; DB 15; Length 972;

Best Local Similarity 46.7%; Pred. No. 2,4e-78; Indels 185; Gaps 14;

Matches 284; Conservative 36; Mismatches 103; Indels 185; Gaps 14;

```

QY 30 LGKKGDTVELTCTASOKSIOF--HMKNNOIKILGNOSFLTCKPSKLNDRADSRSL- 86
DB 537 LVKPSSETLITCTVSGSLTNYGVHWROSPGKLEIGVKNMGSGTEYMAAFISRLTIS 596
QY 87 --WDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOLLVGLTANSDTHLQ--GQSLTLT 142
DB 597 KQTSKNQVSLKLNLSLTADTAVYYC-----ARNDRYAMDYWGQGTLLV 639
QY 143 LESPPGSSPVOCRSRGRKNIQGGKTLVSQLELDSDGTCTVNLQNKVERKIDIVL 202
DB 640 VSSASTKGPSVFPPLAPSSKTSQG--TVALGCL----- 670
QY 203 AFQKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 262
DB 671 -----VKQYFPEPVTVS-----MNSGALTSG----- 691
QY 263 VKRVTQDPLQMGKQLPLHLTLPOALPOYAGSGNLTALBAKTKGLHQEVNLVWMRATOL 322
DB 692 -----VH--TTPAVL--QSSGLYSLSSVTVVPSSSLGTQYI----- 724
QY 323 QKNLTCEWGPSTPKMLSLKLENKEAKVSKREKPPVVLNPEAGMOCCLSDSGQVLLS 382
DB 725 -----CNV-----NHRPSNTKVDKVV----- 740
QY 383 NIKVLPTWSTPVPCEPAPEPKSCDKTHTC-----PELLGSPVFLFPKPKDTLMISRTPE 437
DB 741 -----EPKSCDKTHTCPPCPAPPELLGSPVFLFPKPKDTLMISRTPE 783
QY 438 VTCVVVDVSHEDPEVKENMYVDGVEVHNAKTKPREEOYNSTRYVSVLTVLHODMNGKE 497
DB 784 VTCVVVDVSHEDPEVKENMYVDGVEVHNAKTKPREEOYNSTRYVSVLTVLHODMNGKE 843
QY 498 YKCVSNKALPAPIETKISKAGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIA 557
DB 844 YKCVSNKALPAPIETKISKAGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIA 903
QY 558 VEMESNGQPENNYKTTTPVLDSDGSFELYSKLTVDKSRMOQGNVFCSVVHREALHNHYTQ 617
DB 904 VEMESNGQPENNYKTTTPVLDSDGSFELYSKLTVDKSRMOQGNVFCSVVHREALHNHYTQ 963
QY 618 KSLSLSPG 625
DB 964 KSLSLSPG 971

```

RESULT 82  
US-10-418-836-39  
; Sequence 39, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:

```

; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: G0741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-39

```

Query Match 36.0%; Score 1230.5; DB 15; Length 975;

Best Local Similarity 46.7%; Pred. No. 2,4e-78; Indels 185; Gaps 14;

Matches 284; Conservative 36; Mismatches 103; Indels 185; Gaps 14;

```

QY 30 LGKKGDTVELTCTASOKSIOF--HMKNNOIKILGNOSFLTCKPSKLNDRADSRSL- 86
DB 540 LVKPSSETLITCTVSGSLTNYGVHWROSPGKLEIGVKNMGSGTEYMAAFISRLTIS 599
QY 87 --WDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOLLVGLTANSDTHLQ--GQSLTLT 142
DB 600 KQTSKNQVSLKLNLSLTADTAVYYC-----ARNDRYAMDYWGQGTLLV 642
QY 143 LESPPGSSPVOCRSRGRKNIQGGKTLVSQLELDSDGTCTVNLQNKVERKIDIVL 202
DB 643 VSSASTKGPSVFPPLAPSSKTSQG--TVALGCL----- 673
QY 203 AFQKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 262
DB 674 -----VKQYFPEPVTVS-----MNSGALTSG----- 694
QY 263 VKRVTQDPLQMGKQLPLHLTLPOALPOYAGSGNLTALBAKTKGLHQEVNLVWMRATOL 322
DB 695 -----VH--TTPAVL--QSSGLYSLSSVTVVPSSSLGTQYI----- 727
QY 323 QKNLTCEWGPSTPKMLSLKLENKEAKVSKREKPPVVLNPEAGMOCCLSDSGQVLLS 382
DB 728 -----CNV-----NHRPSNTKVDKVV----- 743
QY 383 NIKVLPTWSTPVPCEPAPEPKSCDKTHTC-----PELLGSPVFLFPKPKDTLMISRTPE 437
DB 744 -----EPKSCDKTHTCPPCPAPPELLGSPVFLFPKPKDTLMISRTPE 786
QY 438 VTCVVVDVSHEDPEVKENMYVDGVEVHNAKTKPREEOYNSTRYVSVLTVLHODMNGKE 497
DB 787 VTCVVVDVSHEDPEVKENMYVDGVEVHNAKTKPREEOYNSTRYVSVLTVLHODMNGKE 846
QY 498 YKCVSNKALPAPIETKISKAGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIA 557
DB 847 YKCVSNKALPAPIETKISKAGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIA 906
QY 558 VEMESNGQPENNYKTTTPVLDSDGSFELYSKLTVDKSRMOQGNVFCSVVHREALHNHYTQ 617
DB 907 VEMESNGQPENNYKTTTPVLDSDGSFELYSKLTVDKSRMOQGNVFCSVVHREALHNHYTQ 966
QY 618 KSLSLSPG 625
DB 967 KSLSLSPG 974

```



```

QY 441 VVVVSHEDPEVCKFMWYVDCVFNHAKTKREBOVNSTRYVSVLTVLHODMLNGKEYKC 500
| | | | |
DB 197 VVVVDSHEDEVEFNNYVDCVFNHAKTKREBOVNSTRYVSVLTVLHODMLNGKEYKC 256
| | | | |
QY 501 KVSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEM 560
| | | | |
DB 257 KVSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEM 316
| | | | |
QY 561 ESNQGEENNYKTPPVLDSDGSFPLYSKLTVDKSRMQQGVNFSQVMEHNLHNYTKSL 620
| | | | |
DB 317 ESNQGEENNYKTPPVLDSDGSFPLYSKLTVDKSRMQQGVNFSQVMEHNLHNYTKSL 376
| | | | |
QY 621 SLSPG 625
| | | | |
DB 377 SLSPG 381

RESULT 86
US-09-773-877A-18
; Sequence 18, Application US/09773877A
; Publication No. US2003017977A1
; GENERAL INFORMATION:
; APPLICANT: Xia, Yu-Ping et al.
; TITLE OF INVENTION: METHODS FOR TREATING INFLAMMATORY SKIN DISEASES
; FILE REFERENCE: REG 710b
; CURRENT APPLICATION NUMBER: US/09/773.877A
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Flc1(2-3)-Fc (Muc3)
US-09-773-877A-18

Query Match 36.0%; Score 1229; DB 12; Length 462;
Best Local Similarity 53.9%; Pred. No. 1,1e-78;
Matches 274; Conservative 36; Mismatches 86; Indels 112; Gaps 18;

QY 178 DSGMTCTVLDQ-----NOKKVEFKIDIVLAFQKASIVYKKEGQVEFSPLAF 227
| | | | |
DB 6 DTGVLCALISCLLTGSSSGRPFVEM-----YSEIPELIHMTGR--ELVIPCVR 55
| | | | |
QY 228 TVERLT-----GSGELMWQAEERASSKSWITFDLKNKEVSVKRVTDQPKQ 273
| | | | |
DB 56 TSPNITVTLKKFPLDLIPDGKRIW-----DSRKGFIISNATYKEIGL----- 99
| | | | |
QY 274 MGKLLPLHLTLFQALPQYASSGNITLLEAKTGKL-HQEVNLV-----MRATOLOKN- 325
| | | | |
DB 100 -----LTCENATV-----NGHL-----YKTNVLTIRQNTIIDVQISTPRPVKLLRGH 141
| | | | |
QY 326 ---LTCENMGPTSPKMLSLKL---ENKEAKVSKR-----EKPVVVLN 362
| | | | |
DB 142 TLVNLCAITPPLNTRVQMTSYPEBKRRASVRRRIQSSHANIFVSLTIDK---MKN 198
| | | | |
QY 363 PEAGMOCCLSDSGQVLESNIKVLPTWSTPVPCEPAPBPKSCDKTHTC-----PELLGGP 417
| | | | |
DB 199 KDKGLYICVRV--SGPSFKSVNTSV--HIYDKAGP---DEPISCDKTHCPCPCAPPELLGGP 253
| | | | |
QY 418 SVFLFPKPKDITLMISTPTPTCVVUVSHEDPEVKNNWVVDGVEVNNATKREBOVNS 477
| | | | |
DB 254 SVFLFPKPKDITLMISTPTPTCVVUVSHEDPEVKNNWVVDGVEVNNATKREBOVNS 313
| | | | |
QY 478 TYRVVSVLTVLHODMLNGKEYKCYSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDEL 537
| | | | |
DB 314 TYRVVSVLTVLHODMLNGKEYKCYSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDEL 373
| | | | |
QY 538 TKQVNSLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMQ 597
| | | | |
DB 374 TKQVNSLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMQ 433
| | | | |

```

```

QY 598 QGNVFCSCVMHEALHNHYTKQSLSPG 625
| | | | |
DB 434 QGNVFCSCVMHEALHNHYTKQSLSPG 461
| | | | |

RESULT 87
US-10-433-108-29
; Sequence 29, Application US/10433108
; Publication No. US20040053370A1
; GENERAL INFORMATION:
; APPLICANT: Eli Lilly and Company
; TITLE OF INVENTION: GLP-1 FUSION PROTEINS
; FILE REFERENCE: X-13991
; CURRENT APPLICATION NUMBER: US/10/433,108
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 60/251,954
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-433-108-29

Query Match 36.0%; Score 1228.5; DB 12; Length 272;
Best Local Similarity 87.5%; Pred. No. 5.9e-79;
Matches 231; Conservative 9; Mismatches 13; Indels 11; Gaps 2;

QY 373 SDSGVLTLESNIKVLPTW-----STPVPCPAPBPKSCDKTHTC-----PELLGGSVFL 421
| | | | |
DB 8 SDSLKQMEEAVALRLFLEWLKNGSPSGAPPSAEPKSCDKTHTCPCPCAPPELLGGSVFL 67
| | | | |
QY 422 FPKPKDITLMISTPTPTCVVUVSHEDPEVKNNWVVDGVEVNNATKREBOVNSTYRV 481
| | | | |
DB 68 FPKPKDITLMISTPTPTCVVUVSHEDPEVKNNWVVDGVEVNNATKREBOVNSTYRV 127
| | | | |
QY 482 VSVLTVLHODMLNGKEYKCYSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ 541
| | | | |
DB 128 VSVLTVLHODMLNGKEYKCYSNNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ 187
| | | | |
QY 542 VSVLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMQGVN 601
| | | | |
DB 188 VSVLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMQGVN 247
| | | | |
QY 602 FSCSVMEHNLHNHYTKQSLSPG 625
| | | | |
DB 248 FSCSVMEHNLHNHYTKQSLSPG 271
| | | | |

RESULT 88
US-10-404-724-23
; Sequence 23, Application US/10404724
; Publication No. US20030203447A1
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Arnold H.
; TITLE OF INVENTION: Methods and Materials for Increasing Expression of Recombinant
; FILE REFERENCE: 13698U501
; CURRENT APPLICATION NUMBER: US/10/404,724
; PRIOR FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/368,530
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-404-724-23

```



Query Match 36.0%; Score 1228.5; DB 12; Length 465;  
Best Local Similarity 45.8%; Pred. No. 1,2e-78;  
Matches 294; Conservative 35; Mismatches 102; Indels 211; Gaps 17;

```
QY 11 LVLVQLALPLPATGKNNKVLG---KKGDVETLTCTASQKKSIOFHMKNNOIKILNOG 66
DB 7 LFLFLMAAQAQAOIOLVQSGPEVKKPGESVKISCKAS---GYFTTKYGMWVWQAPQOG 63
QY 67 -----SFLTKGPSKUNDRADSRSLMDQGNF-----LIIKLIKEDSDTYI 108
DB 64 LKMGWINTYEEPEYGGD-----FKGRFTTLDSTSTAYLEISSLRSDTYAF 114
QY 109 CEVEDQKEVQLVFGGLTANSDTHLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKT 168
DB 115 C-----ARFGSAVD-----YMGCTLVTVSSASTKGPSVFLPAPSSTSGG-T 157
QY 169 LVSQLELDQSGTWTCTVLQNKVEFKIDIVLAFQKASSIYVKKEGQVEFSPLAFT 228
DB 158 AALGCL-----WNSGALITSG-----VKDYFPEPYT 173
QY 229 VEKLTGSGELMWQABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOAL 288
DB 174 VS-----WNSGALITSG-----VH-TFPAVL 192
QY 289 PoyAGSNLTLALFAKTGKLHQBENLVVWRATOLQKULTCEWGPSPKMLSLKLENKE 348
DB 193 -QSGGLYSLSSVTVVPSSTSGTQTYI-----CNV-----NHK 223
QY 349 AKYSKREKRPVWVLPNPAAGMOCLLSDSGVLLSNIKVLPTWSTPVPCEAPPEKSCDTH 408
DB 224 PSNTKVKRV-----EPKSCDTH 242
QY 409 TC-----PELLGSPVFLPPPKDITMISTPEVTCVYVDVSHEDPEVKFMYVDGVEV 463
DB 243 TCPCEAPPELLGSPVFLPPPKDITMISTPEVTCVYVDVSHEDPEVKFMYVDGVEV 302
QY 464 HNAKTKREEOYNSTYRVSVTLVLDHODMLNGKEYCKCKSNKALPAPIKTIKAKGPR 523
DB 303 HNAKTKREEOYNSTYRVSVTLVLDHODMLNGKEYCKCKSNKALPAPIKTIKAKGPR 362
QY 524 EPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVEMESNGQPENNYKTPPVLDSDGSF 583
DB 363 EPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVEMESNGQPENNYKTPPVLDSDGSF 422
QY 584 FLVSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSPG 625
DB 423 FLVSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSPG 464
```

## RESULT 89

US-10-656-769-32  
Sequence 32, Application US/10656769  
Publication No. US2004009712A1  
GENERAL INFORMATION:  
APPLICANT: Varnum, Brian  
APPLICANT: Witte, Alison  
APPLICANT: Vezina, Chris  
APPLICANT: Wong, Lu Min  
TITLE OF INVENTION: Olan, Xuejing  
FILE REFERENCE: 01,1554  
CURRENT APPLICATION NUMBER: US/10/656,769  
CURRENT FILING DATE: 2003-09-05  
NUMBER OF SEQ ID NOS: 79  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 32  
LENGTH: 467  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-656-769-32

Query Match 36.0%; Score 1228; DB 16; Length 467;

Best Local Similarity 45.3%; Pred. No. 1,3e-78;  
Matches 290; Conservative 39; Mismatches 101; Indels 210; Gaps 18;

```
QY 16 LALLPAAATGQ-----NKVVLGKGDVETLTCTASQKKSIOFHM-----KNSN 57
DB 7 LALLLAVLQGVCAEVLQMGABEVKKPGESVKISCKAS---GYSFSEFWWAWRWQAPQOG 65
QY 58 QIKILNGSGFLTKGPS---KLANDRARSRL-----WDQGNFPLIINKLIKEDSDTYI 110
DB 66 WMGIIHFGASDTRKSPFGQVTTISADNSNATYLOW-----SLKASDTAMFCA 116
QY 111 VEDQKEVQLVFGGLTANSDTHLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKT 170
DB 117 ---RQRELDYFDY-----WGGCTLVTVSSASTKGPSVFLPAPSSTSGG-TAA 161
QY 171 VSQLELDQSGTWTCTVLQNKVEFKIDIVLAFQKASSIYVKKEGQVEFSPLAFTVE 230
DB 162 LGCL-----WNSGALITSG-----VKDYFPEPYT 177
QY 231 KLTVSGELMWQABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOALPQ 290
DB 178 -----WNSGALITSG-----VH-TFPAVL-Q 195
QY 291 YAGSGNLTLALFAKTGKLHQBENLVVWRATOLQKULTCEWGPSPKMLSLKLENKE 350
DB 196 SSGLYSLSSVTVVPSSTSGTQTYI-----CNV-----NHKPS 227
QY 351 VSKREKRPVWVLPNPAAGMOCLLSDSGVLLSNIKVLPTWSTPVPCEAPPEKSCDTHTC 410
DB 228 NTKVDKRV-----EPKSCDTHTC 246
QY 411 -----PELLGSPVFLPPPKDITMISTPEVTCVYVDVSHEDPEVKFMYVDGVEV 465
DB 247 PCEPAPPELLGSPVFLPPPKDITMISTPEVTCVYVDVSHEDPEVKFMYVDGVEV 306
QY 466 AKTKREEOYNSTYRVSVTLVLDHODMLNGKEYCKCKSNKALPAPIKTIKAKGPR 525
DB 307 AKTKREEOYNSTYRVSVTLVLDHODMLNGKEYCKCKSNKALPAPIKTIKAKGPR 366
QY 526 QVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVEMESNGQPENNYKTPPVLDSDGSF 585
DB 367 QVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVEMESNGQPENNYKTPPVLDSDGSF 426
QY 586 YSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSPG 625
DB 427 YSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSPG 466
```

## RESULT 90

US-09-747-669-3  
Sequence 3, Application US/09747669  
Patent No. US20020122807A1  
GENERAL INFORMATION:  
APPLICANT: Dan, Michael D.  
APPLICANT: Saleh, Mansoor  
TITLE OF INVENTION: ANTIGEN BINDING FRAGMENTS, DESIGNATED  
TITLE OF INVENTION: 4BS THAT SPECIFICALLY DETECT CANCER CELLS, NUCLEOTIDES  
TITLE OF INVENTION: ENCODING THE FRAGMENTS, AND USE THEREOF FOR THE PROPHYLAXIS  
FILE REFERENCE: 316082001001  
CURRENT APPLICATION NUMBER: US/09/747,669  
CURRENT FILING DATE: 2002-04-08  
PRIORITY APPLICATION NUMBER: US 09/111,286  
PRIORITY FILING DATE: 1998-07-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic construct  
US-09-747-669-3



ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic construct  
US-10-433-108-21

Query Match 36.0%; Score 1227.5; DB 12; Length 272;  
Best Local Similarity 88.3%; Pred. No. 6.9e-79;  
Matches 233; Conservative 4; Mismatches 16; Indels 11; Gaps 2;

QY 373 SDSQVLLSNIKVLPTM-----STVPCAPAPKSCDKTHTC-----PELLGSPSVFL 421  
DB 8 SDVSSYLEEQAKERIALVLKRGSSGAPPPSABPKSCDKTHTCPPCAPPELLGSPSVFL 67  
QY 422 PPPKCDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVYDGEVHNAAKTRPREOYNSTRYV 481  
DB 68 PPPKCDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVYDGEVHNAAKTRPREOYNSTRYV 127  
QY 482 VSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 541  
DB 128 VSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 187  
QY 542 VSLTCLVKGFYPSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV 601  
DB 188 VSLTCLVKGFYPSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV 247  
QY 602 FSCSVMEALHNHYTOKSLSLSPG 625  
DB 248 FSCSVMEALHNHYTOKSLSLSPG 271

## RESULT 93

US-10-378-567-2  
Sequence 2, Application US/10378567  
Publication No. US20040006208A1  
GENERAL INFORMATION:  
APPLICANT: KARPUSAS, MICHAEL  
APPLICANT: HSU, YEN-MING  
APPLICANT: TAYLOR, FREDERICK R.  
APPLICANT: ZHENG, ZHONGLI  
TITLE OF INVENTION: CO-CRYSTAL STRUCTURE OF MONOCLONAL ANTIBODY 5C8 AND  
TITLE OF INVENTION: CD154, AND USE THEREOF IN DRUG DESIGN  
FILE REFERENCE: A096CON1  
CURRENT APPLICATION NUMBER: US/10/378.567  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/27352  
PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: 60/276,452  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/229,933  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 448  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: humanized 5c8 heavy chain amino acid  
US-10-378-567-2

Query Match 36.0%; Score 1227.5; DB 15; Length 448;  
Best Local Similarity 46.2%; Pred. No. 1.4e-78;  
Matches 285; Conservative 42; Mismatches 97; Indels 193; Gaps 17;  
QY 25 GKNVVLGKGPDTVELTCTASQK--KSIOFHKNNSQIKILGNQG--SFL-----TKGPSKL 76  
DB 8 GAELV--KPGASVLTSCASGIFTSYMYM-----VKQAPGQGLEMIIGELINPENGDTNF 60  
QY 77 NDRADRSRLW---DQGNFLLIKNLKIEDSTYICEVEDQKEEVQLLVFGLNANSDTHL 133  
DB 61 NEKFKSKATLVTDKASATAWELSLRSEDTAVYVYCTRSRDRNDM----- 106

QY 134 LOGSLTLTLSPSSPSVQCRRPKNIGQGKTLVSQLELDQSGTWCTVLQONQKV 193  
DB 107 SMGGTLVTVSSASTKPSVFPPLAPSSKTSQSG--TALAGCL----- 146  
QY 194 EFKIDIVLAFQKASSIVYKKEGQVEPFPPLAFTVEKLTGSGELMWQABASSKSMIT 253  
DB 147 -----VQDYFPEPVTGS-----NNSGALTSQ----- 167  
QY 254 FDLKNKESVYKRVYQDPKLGKGLPLHLTLPLQALPYAGSGNLTALAEATGKLGHOEVN 313  
DB 168 -----VH-TFPAVL--QSSGLYSLSVTVYVSSLSLGTQTY 199  
QY 314 LVVVRATOLQKNTLCEVWGFTSPKLMLSLKENKAYSKREKPVVVLNPEAGWQCILS 373  
DB 200 I-----CNY-----NHKPSNTKYDKV----- 216  
QY 374 DSQGVLLSNIKVLPTMSTVPCAPAPKSCDKTHTC-----PELLGSPSVFLPPPKD 428  
DB 217 -----EPKSCDKTHTCPPCAPPELLGSPSVFLPPPKD 250  
QY 429 TLMISRTPEVTCVVVDVSHEDPEVKFNNYVYDGEVHNAAKTRPREOYNSTRYVSVTLVL 488  
DB 251 TLMISRTPEVTCVVVDVSHEDPEVKFNNYVYDGEVHNAAKTRPREOYNSTRYVSVTLVL 310  
QY 489 HODVLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLV 548  
DB 311 HODVLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLV 370  
QY 549 KGFPYSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVSCVMH 608  
DB 371 KGFPYSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVSCVMH 430  
QY 609 EALHNHYTOKSLSLSPG 625  
DB 431 EALHNHYTOKSLSLSPG 447

## RESULT 94

US-10-108-260A-4285  
Sequence 4285, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1 full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108.260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4285  
LENGTH: 471  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-108-260A-4285

Query Match 35.9%; Score 1226; DB 15; Length 471;  
Best Local Similarity 51.2%; Pred. No. 1.8e-78;  
Matches 281; Conservative 18; Mismatches 72; Indels 178; Gaps 15;

QY 163 IOGKTLVSQLELDQSGTWCTVLQONQKVEFKIDIVLAFQKASSIVYKKEGQVEFS 222  
DB 14 ISSGQ-----SQVPLVQSGT-----EYKKGASVNIIS 40  
QY 223 FP--LAFT-----VEKLTGSGELMWQABASSKSMITFDLKNKESVYKRVYQDPKQ 273  
DB 41 CKAPGYFTTFYFMHWVWQAPGQGLEWMIIRINPSSGRS-----SVSQK----- 82  
QY 274 MGKGLPLHLTLPLQALPYAGSGNLTALAEATGKLGHOEVNLVVVRATOLQKNTLCE 329  
DB 83 -----FEGRLTLTADISTTANHEL-----RNLTSDDTGV 112  
QY 330 -----VMGPTSPKLMLSLKENKAYSKREKPVVVLNPEA----- 365

Db 113 YCTTRMKVNVKGEDNMGQSLVIVSS-----ASTKGSVFLPAPSSTKSTSGT 163  
Qy 366 GMMQCLISD-----SGOVLLSNIKVLPT-----TWSTPVPCPAP----- 399  
Db 164 AALGCLVXKDYFPEPVTVMWNSG--ALTSVHTFPAVLQSSGLSYLSLSSVTVVSSLSGTOT 221  
Qy 400 -----EPKSCDKHTTC-----PELLGSPVFLPPEPKDTLMTSRP 436  
Db 222 YICNVNHNKSNNTVDEKVEPKSCDKHTTCPCPAPPELLGSPVFLPPEPKDTLMTSRP 281  
Qy 437 EYTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYNSTYRVSVLTVLHQMNLNK 496  
Db 282 EYTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYNSTYRVSVLTVLHQMNLNK 341  
Qy 497 EYTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYNSTYRVSVLTVLHQMNLNK 556  
Db 342 EYTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYNSTYRVSVLTVLHQMNLNK 401  
Qy 557 AVEWESNGOPENNYKTTTPVLVDSGSPFLYSKLTVDKSRMGOGNVSCSVMEHALNHYT 616  
Db 402 AVEWESNGOPENNYKTTTPVLVDSGSPFLYSKLTVDKSRMGOGNVSCSVMEHALNHYT 461  
Qy 617 QKSLSLSPG 625  
Db 462 QKSLSLSPG 470

RESULT 95  
US-10-108-260A-4292  
; Sequence 4292, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 4292  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4292

Query Match 35.9%; Score 1225.5; DB 15; Length 470;  
Best Local Similarity 46.3%; Pred. No. 2e-78;  
Matches 291; Conservative 42; Mismatches 100; Indels 195; Gaps 18;  
Qy 15 QLALLPAAQGNKVVIGKGGDTVELTCTAS--QKKSIOFHMKK-----SNQKIL 62  
Db 20 QVQVQVSGTE---VKKPSSVAVSKAGSGSFSTVFTWROAPGEGLEMMGSIIPIL 74  
Qy 63 GNGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQLLV 122  
Db 75 G-RPNVQKQODRTIASADSSSI---VYMDLRLTIEDTAIFYCAI-----LLE 120  
Qy 123 FGLTANSDTLLQGSITLTLLESPGSSPSVQCRSPGKNIQGGKITLSVQLELDGSGTW 182  
Db 121 HEVVALFD-HMGQGTILVTVSSASTKG--PSVFLPAPSSTSGG-TAALGCL----- 168  
Qy 183 TCVTLLQKQKVEFRIDIVLAFQKASIVYKKEGEQVEFSPLAFTEKLTGSGELMWQA 242  
Db 169 -----VKDYFPEPVTS-----WNS 183  
Qy 243 ERASSSKSWITFDLKNKEVSVKVTQDPKLOMGKPLHLTLPOALPOVAGSGLTLALE 302  
Db 184 GALTSG-----VH-TFPAVL-QSSGIVSLSSVVT 210  
Qy 303 AKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKAIVSKREKPVWLN 362  
Db 211 VPSSISLGTQTYI-----CNV-----NHKPSNTKVDKAV----- 238

Qy 363 PEAGMMQCLISDSGOVLLSNIKVLPTWSTPVPCPAPPEPKSCDKHTTC-----PELLGSP 417  
Db 239 -----EPKSCDKHTTCPCPAPPELLGSP 261  
Qy 418 SVFLPPEPKDTLMTSRPEVTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYN 477  
Db 262 SVFLPPEPKDTLMTSRPEVTCVAVDVSHEDPEVKFNMYVDGVEVHNAKTPREEOYN 321  
Qy 478 TYRVSVLTVLHQMNLNKGEYCKVSNKALPAPIKTIKSKAGOPREPQVTLPPSRDEL 537  
Db 322 TYRVSVLTVLHQMNLNKGEYCKVSNKALPAPIKTIKSKAGOPREPQVTLPPSRDEL 381  
Qy 538 TKQVSLTCLVAGFYPSDIAVWESNGOPENNYKTTTPVLVDSGSPFLYSKLTVDKSRM 597  
Db 382 TKQVSLTCLVAGFYPSDIAVWESNGOPENNYKTTTPVLVDSGSPFLYSKLTVDKSRM 441  
Qy 598 QGNVSCSVMEHALNHYTQKSLSLSPG 625  
Db 442 QGNVSCSVMEHALNHYTQKSLSLSPG 469

RESULT 96  
US-10-282-162-34  
; Sequence 34, Application US/10282162  
; Publication No. US20030143697A1  
; GENERAL INFORMATION:  
; APPLICANT: REGENERON PHARMACEUTICALS, INC.  
; TITLE OF INVENTION: RECEPTOR BASED ANTAGONISTS, AND METHODS OF MAKING  
; FILE REFERENCE: REG 203-B-US  
; CURRENT APPLICATION NUMBER: US/10/282,162  
; CURRENT FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: 09/787,835  
; PRIOR FILING DATE: 1999-09-22  
; PRIOR APPLICATION NUMBER: PCT/US99/22045  
; PRIOR FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 900  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-282-162-34

Query Match 35.9%; Score 1225.5; DB 14; Length 900;  
Best Local Similarity 48.3%; Pred. No. 4.8e-78;  
Matches 284; Conservative 45; Mismatches 112; Indels 147; Gaps 19;  
Qy 86 LW-----DOGNFPLIINKLIEDSDTYICEVEDQKEVQL-LVFGLTANSDTLL--Q 135  
Db 411 LMRPFLTLDNTGNYTOMLN-----TTCCKVAFLEVVQKDSCTNSPMKLPVHLXYIE 464  
Qy 136 GOSLTLTLESPGSSPS-----VQCRSPGKN-IQGGKITLSVQLELDGSGTWC 184  
Db 465 YGIRITCPNVQGFPSVSVKPTITWYMGCYKQFNFNVIPEGNLSFLIALISNNGYTC 524  
Qy 185 TVLQNKQKVE---KIDIVLAFQKAS-----SVYKKE-GEQ-----VEFSF 223  
Db 525 VVYTPENGRTFLTRLTVKVGSPKNAVAPVHISNDHVYKEKEGEBLLIPCTYFESF 584  
Qy 224 PLAFTEKLTGSGELMWQERASSSKSWITFDLK-NKEVSVKVTQDPKLOMGKPLHL 282  
Db 585 -----LNDNRBVMWITD--GKKDDITIDVTINESISHSTBEDTRQI----- 627  
Qy 283 TLPOALPOVAGSGLTLALEAKTKLHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSL 342  
Db 628 -----LSIKVTSB-----DLKRSYVCHA----- 646  
Qy 343 KLENKAIVSKREKPVWNLNPEAGMMQCLISDSGOVLLSNIKVLPTWSTPVPCPAPPEK 402  
Db 647 ---RSK-----GEVAKAKVK-----OKVAPRYTVE 671  
Qy 403 SCDKHTTC-----PELLGSPVFLPPEPKDTLMTSRPEVTCVAVDVSHEDPEVKFNMY 457

Db 672 SGDTHHTPCPAPABELLGGPSVFLPPPKDITLMSRPEVTCVVVDVSHEDPEVKFMY 731  
Qy 458 VDGVEVHNAKTKRPREQVNSTYRVVSVLTVTHQDMLNKEVKCKSNALPAPIEKTISK 517  
Db 732 VDGVEVHNAKTKRPREQVNSTYRVVSVLTVTHQDMLNKEVKCKSNALPAPIEKTISK 791  
Qy 518 AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTPPV 577  
Db 792 AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTPPV 851  
Qy 578 DSDGSFFLYSKLTVDKSRMQGNVSCSVMEHALNHYTOKSLSPG 625  
Db 852 DSDGSFFLYSKLTVDKSRMQGNVSCSVMEHALNHYTOKSLSPG 899

RESULT 97  
US-10-683-255-12  
; Sequence 12, Application US/10683255  
; Publication No. US20040063910A1  
; GENERAL INFORMATION:  
; APPLICANT: Kavanaugh, William M.  
; APPLICANT: Ballinger, Marcus  
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR  
; TITLE OF INVENTION: RECEPTOR-IMMUNOGLOBULIN FUSION  
; FILE REFERENCE: PP01474.101  
; CURRENT APPLICATION NUMBER: US/10/683.255  
; CURRENT FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: 09/499,846  
; PRIOR FILING DATE: 2000-02-07  
; PRIOR APPLICATION NUMBER: 60/119,002  
; PRIOR FILING DATE: 1999-02-08  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 488  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-683-255-12

Query Match 35.9%; Score 1225; DB 12; Length 488;  
Best Local Similarity 67.3%; Pred. No. 2.3e-78;  
Matches 253; Conservative 18; Mismatches 56; Indels 49; Gaps 8;  
Qy 293 GSGVLTALAEKTKGLH---EVNLVYMRATQLOKLT---TCEWGFSPKMLSLKE-- 345  
Db 118 GSIHRTYQLDVERSPHPILOAGLPANKTVALGSNVEFMCKVYSDPPIQWLKHIEVN 177  
Qy 346 -----NKEAKVSKREKPVVVLN-----PEAGMOCILSDSGQVLLSENIK 385  
Db 178 GSKTGPDLPIYVQLKTAGVTTDKEMEVHLHKNVSFEADAGEYTCLAGNSIGLSHNS--- 234  
Qy 386 VLPWTST-----PVPCEPAP---EPKSCDKHTC-----PELLGGSVFLPPPKPD 429  
Db 235 ---AMLTVLALERPAVMTSPVLEPSSCKHTHCPCPAPBELGGSVFLPPPKPD 291  
Qy 430 LMIKRTPEVTCVVVDVSHEDPEVKFNKYVDGEVHNAKTKRREQVNSTYRVVSVLT 489  
Db 292 LMIKRTPEVTCVVVDVSHEDPEVKFNKYVDGEVHNAKTKRREQVNSTYRVVSVLT 351  
Qy 490 QDMNNGKREYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL 549  
Db 352 QDMNNGKREYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL 411  
Qy 550 GFYPSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCSVME 609  
Db 412 GFYPSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCSVME 471  
Qy 610 ALAHNHYTOKSLSPG 625  
Db 472 ALAHNHYTOKSLSPG 487

RESULT 98  
US-10-193-616-14  
; Sequence 14, Application US/10193616  
; Publication No. US20030096355A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhang, Ke  
; TITLE OF INVENTION: Isolation, Identification, and Characterization of  
; TITLE OF INVENTION: Ymk5, a novel  
; TITLE OF INVENTION: member of the TNF-Receptor Supergene Family  
; FILE REFERENCE: 01017/35551A  
; CURRENT APPLICATION NUMBER: US/10/193.616  
; CURRENT FILING DATE: 2002-07-11  
; PRIOR APPLICATION NUMBER: US/09/611,989  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/143,137  
; PRIOR FILING DATE: 1999-07-07  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; OTHER INFORMATION: ymk5-Fc fusion protein  
US-10-193-616-14

Query Match 35.9%; Score 1224.5; DB 14; Length 396;  
Best Local Similarity 92.1%; Pred. No. 1.9e-78;  
Matches 233; Conservative 1; Mismatches 14; Indels 5; Gaps 1;  
Qy 378 VLESNIKVLPTWSTPVPCPAPBELGGSVFLPPPKPDITMI 432  
Db 143 VLQECNSTANTVYSSSNAAEPKSCDKHTHCPCPAPBELGGSVFLPPPKPDITMI 202  
Qy 433 SRTEPVTCVVVDVSHEDPEVKFNKYVDGEVHNAKTKRREQVNSTYRVVSVLT 492  
Db 203 SRTEPVTCVVVDVSHEDPEVKFNKYVDGEVHNAKTKRREQVNSTYRVVSVLT 262  
Qy 493 LNKREYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL 552  
Db 263 LNKREYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL 322  
Qy 553 PSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCSVME 612  
Db 323 PSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCSVME 382  
Qy 613 NHYTOKSLSPG 625  
Db 383 NHYTOKSLSPG 395

RESULT 99  
US-10-471-151-32  
; Sequence 32, Application US/10471151  
; Publication No. US20040086908A1  
; GENERAL INFORMATION:  
; APPLICANT: Chandrasekhar, Yasmin A.  
; APPLICANT: Novak, Julia E.  
; APPLICANT: Foster, Donald C.  
; APPLICANT: Menfeng, Xu  
; APPLICANT: Jaspers, Stephen R.  
; TITLE OF INVENTION: Soluble Heterodimeric Cytokine Receptor  
; FILE REFERENCE: 01-10PC  
; CURRENT APPLICATION NUMBER: US/10/471.151  
; CURRENT FILING DATE: 2003-09-08  
; PRIOR APPLICATION NUMBER: 60/274,560  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/299,865  
; PRIOR FILING DATE: 2001-06-21  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 541

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-471-151-32

Query Match 35.9%; Score 1224.5; DB 16; Length 541;  
Best Local Similarity 53.8%; Pred. No. 2.8e-78;  
Matches 278; Conservative 26; Mismatches 100; Indels 113; Gaps 13;

```

QY 209 SIYKKEGEO-----VEFSPLAFTYEKLTGSGELMMQARASSSKSWITFDLK 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 37 SIEYKTYGERDVAKKGCORITRKSCNLTVEGTNLT---ELYARVTAASAGRSATKMT 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 258 NKEVSVKRYTODP-----KLQMGKULPLHLTPQALPOYAGSGNLTALAEKTK 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 DRPSSLGHTTLKRPDYTCISKVRSIQM-----IVHPTPTIRAGDGH-RLTLEIDIFHD 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 LHOENVLVVVRATOLQK-----LTCEVWGPTSPKMLSLK-L 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 LFYHLELVNRTYQVHNLGKQREYEFGLTPDTEFLGTIMICVPTAKESAPYMCVKTL 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 345 ENKEAKVSKREKPVWVNLPEA-----GMQCLLSD-----SGQVLLBSNIVK 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 PDRTWTASTKGPSVFPLAPSSKSTSGGTALGCLVXDYFPEPVTVSMNSG--ALTSGVHT 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 LP-----TWSTPVPCEPAP-----EPKSCDKTHTC--- 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 264 FPAVLQSSGLYSLSVTVTPSSSLGTQTYICNVNHNKPSNTKVDKKEPSPSCDKTHTCPPC 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 411 --PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 324 PABELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 KPREQYNSTYRVSVLTTLVHODMNLGKEKKCKVSNKALPAPIEKTIISKAKQPREPQVY 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 KPREQYNSTYRVSVLTTLVHODMNLGKEKKCKVSNKALPAPIEKTIISKAKQPREPQVY 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 TLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQPENNYKTTTPVLDSDGSFPLYSK 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 TLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQPENNYKTTTPVLDSDGSFPLYSK 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 LTVDKSRMQQGNVFSCSVMEALHNHYTQKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 504 LTVDKSRMQQGNVFSCSVMEALHNHYTQKSLSLSPG 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 100  
US-10-471-151-31  
Sequence 31, Application US/10471151  
Publication No. US20040086908A1  
GENERAL INFORMATION:  
APPLICANT: Chandrasekhar, Yashmin A.  
APPLICANT: Novak, Julia E.  
APPLICANT: Foster, Donald C.  
APPLICANT: Wenfeng, Xu  
APPLICANT: Jaspers, Stephen R.  
TITLE OF INVENTION: Soluble Heterodimeric Cytokine Receptor  
FILE REFERENCE: 01-10PC  
CURRENT APPLICATION NUMBER: US/10/471.151  
CURRENT FILING DATE: 2003-09-08  
PRIOR APPLICATION NUMBER: 60/274,560  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/299,865  
PRIOR FILING DATE: 2001-06-21  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 31  
LENGTH: 558  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-471-151-31

Query Match 35.9%; Score 1224.5; DB 16; Length 558;  
Best Local Similarity 53.8%; Pred. No. 3e-78;

```

Matches 278; Conservative 26; Mismatches 100; Indels 113; Gaps 13;
QY 209 SIYKKEGEO-----VEFSPLAFTYEKLTGSGELMMQARASSSKSWITFDLK 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 54 SIEYKTYGERDVAKKGCORITRKSCNLTVEGTNLT---ELYARVTAASAGRSATKMT 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 258 NKEVSVKRYTODP-----KLQMGKULPLHLTPQALPOYAGSGNLTALAEKTK 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 DRPSSLGHTTLKRPDYTCISKVRSIQM-----IVHPTPTIRAGDGH-RLTLEIDIFHD 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 LHOENVLVVVRATOLQK-----LTCEVWGPTSPKMLSLK-L 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 LFYHLELVNRTYQVHNLGKQREYEFGLTPDTEFLGTIMICVPTAKESAPYMCVKTL 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 345 ENKEAKVSKREKPVWVNLPEA-----GMQCLLSD-----SGQVLLBSNIVK 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 PDRTWTASTKGPSVFPLAPSSKSTSGGTALGCLVXDYFPEPVTVSMNSG--ALTSGVHT 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 LP-----TWSTPVPCEPAP-----EPKSCDKTHTC--- 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 FPAVLQSSGLYSLSVTVTPSSSLGTQTYICNVNHNKPSNTKVDKKEPSPSCDKTHTCPPC 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 411 --PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 PABELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKT 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 KPREQYNSTYRVSVLTTLVHODMNLGKEKKCKVSNKALPAPIEKTIISKAKQPREPQVY 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 401 KPREQYNSTYRVSVLTTLVHODMNLGKEKKCKVSNKALPAPIEKTIISKAKQPREPQVY 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 TLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQPENNYKTTTPVLDSDGSFPLYSK 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 461 TLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQPENNYKTTTPVLDSDGSFPLYSK 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 LTVDKSRMQQGNVFSCSVMEALHNHYTQKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 521 LTVDKSRMQQGNVFSCSVMEALHNHYTQKSLSLSPG 557
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: August 3, 2004, 13:48:02  
Job time: 67.8475 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 5.92937 Seconds

(without alignments)  
4185.504 Million cell updates/sec

Title: SEQ6

Perfect score: 1317

Sequence: 1 MNREVPRHLLVQLALP.....VISFLGIGVACVLRTR 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0 %  
Maximum Match 100 %

Listing first 125 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023	77.7	458	1 RWHUT4	T-cell surface gly
2	875	66.4	432	1 RWCZT4	T-cell surface gly
3	790	60.0	432	1 RWCZT4	T-cell surface gly
4	596.5	45.3	459	2 A46254	CD4 precursor - ra
5	562	42.7	432	2 S30193	T-cell surface gly
6	480.5	37.2	457	2 A27449	T-cell surface gly
7	475	36.1	457	1 RWMST4	T-cell surface gly
8	357	27.1	71	1 160082	CD4 receptor - hum
9	332	25.2	240	2 A39016	T-cell surface gly
10	305.5	23.2	99	2 S21461	T-cell surface gly
11	280.5	21.3	99	2 S21462	T-cell surface gly
12	156.5	11.9	538	2 JC2457	vascular cell adhe
13	133	10.1	1259	2 S36126	neural cell adhesi
14	131	9.9	1260	1 S05479	neural cell adhesi
15	129	9.8	338	2 JC4776	limbic-system-asso
16	124.5	9.5	584	2 150419	8-glycerin precurs
17	124	9.4	739	2 J06075	vascular cell adhe
18	123	9.3	338	2 JC5519	50K glycoprotein p
19	120	9.1	1091	2 A58532	glial cell membran
20	120	9.1	1277	2 T30532	neural cell adhesi
21	120	9.0	2222	2 T13924	sdh protein - frui
22	118.5	9.0	584	2 T08678	hypothetical prote
23	118	9.0	1091	2 S01998	neural cell adhesi
24	118	9.0	1197	2 T30581	protein unc-22 (im
25	118	9.0	6831	2 A88852	tyrosin (samlari
26	118	9.0	6839	2 S57242	hypothetical prote
27	118	9.0	7160	2 T27935	oploid-binding pro
28	117.5	8.9	345	2 S03199	vascular cell adhe
29	117	8.9	739	2 A41288	

30	117	8.9	1011	2 T13669	neuromusculin - fr
31	116.5	8.8	345	2 JC4025	oploid-binding cel
32	115	8.7	647	2 B41288	vascular cell adhe
33	113	8.6	1018	2 JC4211	neural adhesion pr
34	112.5	8.5	120	2 S46374	secretory componen
35	111.5	8.5	173	2 A6374	ig kappa chain V-J
36	111	8.4	702	2 A36319	carcinoembryonic a
37	111	8.4	1091	2 S33850	fibronectin-binding
38	110.5	8.4	398	2 T49443	gene 2B4 protein -
39	110.5	8.4	2029	1 TDFLTK	protein-tyrosine-p
40	110	8.4	521	2 S34338	biliary glycoprote
41	110	8.4	739	2 JN0581	vascular cell adhe
42	109.5	8.3	279	2 S04693	T-cell receptor de
43	109.5	8.3	304	2 S04663	T-cell receptor ga
44	109	8.3	1257	1 A41060	neural cell adhesi
45	109	8.3	1367	2 A41228	protein-tyrosine k
46	108.5	8.2	333	2 A31923	amalgam protein pr
47	108	8.2	257	2 S00682	IGF Fc receptor al
48	107.5	8.2	122	2 S40370	ig kappa chain - h
49	107	8.1	761	1 TJHUNG	neural cell adhesi
50	107	8.1	3707	2 S18252	heparan sulfate pr
51	107	8.1	4391	2 A38096	perlecan precursor
52	106.5	8.1	111	2 B37266	ig kappa chain V r
53	106.5	8.1	111	2 I38740	ig kappa chain V r
54	106.5	8.1	345	2 JC1239	oploid-binding pro
55	106.5	8.1	2783	2 T34416	hypothetical prote
56	106	8.0	1232	2 T43027	neural cell adhesi
57	106	8.0	1896	2 T08851	Down syndrome cell
58	105.5	8.0	210	2 I49294	CD7 antigen - mous
59	105.5	8.0	1323	2 P00568	connectin 3B - chi
60	105	8.0	2629	2 T32735	telomerase-assoia
61	104.5	7.9	1239	1 A32579	neuroglian - fruit
62	104	7.9	338	2 JC1238	oploid-binding pro
63	104	7.9	344	2 I56551	neurotrilin - rat
64	104	7.9	458	2 JC1509	biliary glycoprote
65	104	7.9	1018	2 A54744	contactin 1 precu
66	104	7.9	1020	2 S05944	neural cell surf
67	104	7.9	1021	2 A57112	neural cell surf
68	103.5	7.9	103	2 S18731	contactin precursor
69	103.5	7.9	108	1 RVMS06	ig kappa chain V r
70	103.5	7.9	117	2 S21668	ig kappa chain V r
71	103.5	7.9	129	1 K1HMK	ig kappa chain pre
72	103.5	7.9	1028	2 A53449	plasmacytoma-asso
73	103.5	7.9	5175	2 T20992	hypothetical prote
74	103.5	7.9	5198	2 T43280	hemocytin precurs
75	103	7.8	519	2 A44783	ecto-ATPase precu
76	102.5	7.8	862	2 I49583	differentiation an
77	102.5	7.8	1348	2 S51656	vascular endotheli
78	102	7.7	458	1 WMMSR1	biliary glycoprote
79	102	7.7	458	2 S23969	cell-adhesion mole
80	102	7.7	458	2 S68177	C-CAM2 protein is
81	102	7.7	521	2 JC1508	biliary glycoprote
82	102	7.7	7962	2 I38346	elastic titin - hu
83	101.5	7.7	98	2 PH1066	ig light chain V r
84	101.5	7.7	725	1 TJMSNG	neural cell adhesi
85	101.5	7.7	858	1 IURMNC	neural cell adhesi
86	101.5	7.7	1115	1 TJMSNL	neural cell adhesi
87	101	7.7	135	2 S10004	hypothetical prote
88	101	7.7	1036	2 S22383	axomin 1 precursor
89	101	7.7	3375	2 T19821	hypothetical prote
90	100.5	7.6	108	1 RVMS61	ig kappa chain V r
91	100.5	7.6	765	2 C42632	cell adhesion mole
92	100.5	7.6	806	2 A35963	protein-tyrosine k
93	100.5	7.6	812	2 B42662	cell adhesion mole
94	100.5	7.6	932	2 A42632	cell adhesion mole
95	100.5	7.6	26926	1 I38344	titin, cardiac mus
96	100	7.6	464	2 C30127	transmembrane carc
97	100	7.6	526	1 A32164	biliary glycoprote
98	99	7.5	230	2 S33161	ig kappa chain - b
99	98.5	7.5	129	2 S40317	ig kappa chain - h
100	98.5	7.5	349	2 A34815	carcinoembryonic a
101	98.5	7.5	1028	2 I58164	BIG-1 protein - ra
102	98	7.4	483	2 T17346	hypothetical prote

neuromusculin - fr  
oploid-binding cel  
vascular cell adhe  
neural adhesion pr  
secretory componen  
ig kappa chain V-J  
carcinoembryonic a  
fibronectin-binding  
gene 2B4 protein -  
protein-tyrosine-p  
biliary glycoprote  
vascular cell adhe  
T-cell receptor de  
T-cell receptor ga  
neural cell adhesi  
protein-tyrosine k  
amalgam protein pr  
IGF Fc receptor al  
ig kappa chain - h  
neural cell adhesi  
heparan sulfate pr  
perlecan precursor  
ig kappa chain V r  
ig kappa chain V r  
JC1239  
oploid-binding pro  
hypothetical prote  
neural cell adhesi  
Down syndrome cell  
CD7 antigen - mous  
connectin 3B - chi  
telomerase-assoia  
neuroglian - fruit  
oploid-binding pro  
neurotrilin - rat  
biliary glycoprote  
contactin 1 precu  
neural cell surf  
neural cell surf  
contactin precursor  
ig kappa chain V-J  
ig kappa chain V r  
ig kappa chain V r  
ig kappa chain pre  
plasmacytoma-asso  
hypothetical prote  
hemocytin precurs  
ecto-ATPase precu  
differentiation an  
vascular endotheli  
biliary glycoprote  
cell-adhesion mole  
C-CAM2 protein is  
biliary glycoprote  
elastic titin - hu  
ig light chain V r  
neural cell adhesi  
neural cell adhesi  
neural cell adhesi  
hypothetical prote  
axomin 1 precursor  
hypothetical prote  
ig kappa chain V r  
cell adhesion mole  
protein-tyrosine k  
cell adhesion mole  
cell adhesion mole  
titin, cardiac mus  
transmembrane carc  
biliary glycoprote  
ig kappa chain - b  
ig kappa chain - h  
carcinoembryonic a  
BIG-1 protein - ra  
hypothetical prote

103	98	7 4	764	1	ORHUGS
104	98	7 4	880	2	B53743
105	97.5	7 4	108	1	KVMS82
106	97.5	7 4	117	2	S11700
107	97.5	7 4	122	2	S40314
108	97.5	7 4	620	2	JH0593
109	97.5	7 4	1333	2	T178875
110	97.5	7 4	6642	2	T29757
111	97	7 4	194	2	T29925
112	97	7 4	470	2	S22080
113	97	7 4	729	2	A56795
114	97	7 4	733	2	I49293
115	97	7 4	822	2	I49289
116	97	7 4	1091	1	IJCXNL
117	97	7 4	1328	2	T23007
118	97	7 4	1330	2	S49010
119	97	7 4	1447	2	A54100
120	96.5	7 3	117	1	K1HUI2
121	96.5	7 3	122	2	S40351
122	96.5	7 3	125	2	S40349
123	96.5	7 3	130	2	PL0113
124	96.5	7 3	323	2	S01895
125	96.5	7 3	365	2	JC7780

## ALIGNMENTS

RESULT 1  
RMHUT4  
T-cell surface glycoprotein CD4 precursor [validated] - human  
N/Alternate names: T-cell surface antigen T4/Les 3  
C/Species: Homo sapiens (man)  
C/Date: 20-May-1986 #sequence, revision 31-Dec-1988 #text, change 20-Apr-2001  
C/Accession: A90872; A32722; A34194; A53287; I54176; I54297; A02109; A30039  
R/Maddon, P.J.; Littman, D.R.; Godfrey, M.; Maddon, D.E.; Chess, L.; Axel, R.  
Cell 42, 93-104, 1985  
A/Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein  
A/Reference number: A90872; MUID:85254946; PMID:2990730  
A/Accession: A90872  
A/Molecule type: mRNA  
A/Residues: 1-25, 'N', 27-458 <MAD>  
A/Experimental source: clone pT4B  
R/Littman, D.R.; Maddon, P.J.; Axel, R.  
Cell 55, 541, 1988  
A/Title: Corrected CD4 sequence.  
A/Reference number: A90907; MUID:89028665; PMID:3263213  
A/Contents: annotation; revision to residue 26  
R/Camerini, D.; Seed, B.  
Cell 60, 747-754, 1990  
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the viral  
A/Reference number: A32722; MUID:90182664; PMID:2107024  
A/Accession: A32722  
A/Status: nucleic acid sequence not shown; not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 26-426 428-458 <CAM>  
R/Carr, S.A.; Hemling, M.E.; Folema-Wasserman, G.; Sweet, R.W.; Annunzio, K.; Barr, J.R.;  
J. Biol. Chem. 264, 21286-21295, 1989  
A/Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor  
A/Reference number: A34194; MUID:90078233; PMID:2592374  
A/Contents: disulfide bonds; carbohydrate-binding sites  
A/Accession: A34194  
A/Molecule type: protein  
A/Residues: 26-394 <CAR>  
R/Lederman, S.; Demartino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A.  
Mol. Immunol. 28, 1171-1181, 1991  
A/Title: A single amino acid substitution in a common African allele of the CD4 molecule  
A/Reference number: A53287; MUID:92072595; PMID:1961196  
A/Accession: A53287  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 250-264, 'W', 266-280 <LED>  
A/Note: sequence extracted from NCBI backbone (NCBIF:68249)

R/Edwards, M.C.; Gibbs, R.A.  
Genomics 14, 590-597, 1992  
A/Title: A human polymorphism resulting from loss of an Alu.  
A/Reference number: I54176; MUID:93052387; PMID:1330888  
A/Accession: I54176  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-72 <RES>  
A/Cross-references: GB:U47924; GB:M6525; GB:U72506; NID:g1633547; PIDN:AA51309.1; PID  
Hum. Immunol. 30, 99-104, 1991  
A/Title: Humans with OX74-epitope deficiency have a single nucleotide base change in the  
A/Reference number: I54297; MUID:91216786; PMID:1708753  
A/Accession: I54297  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-264, 'W', 266-458 <RE2>  
A/Cross-references: GB:M35160; NID:g179143; PIDN:AA16069.1; PID:g179144  
C/Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) receptor  
C/Genetic: GDB:CD4  
A/Status: GDB:CD4  
A/Cross-references: GDB:119767; OMIM:186940  
A/Map position: 12pter-12p12  
A/Intons: 16/3  
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C/Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein  
F/1-25/Domain: signal sequence #status predicted <Sig>  
F/26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>  
F/34-111/Domain: immunoglobulin homology <IM1>  
F/136-186/Domain: immunoglobulin homology <IM2>  
F/216-299/Domain: immunoglobulin homology <IM3>  
F/321-372/Domain: immunoglobulin homology <IM4>  
F/397-420/Domain: transmembrane #status predicted <TM>  
F/421-458/Domain: intracellular #status predicted <INT>  
F/421-109, 155-184, 328-370/disulfide bonds: #status experimental  
F/296, 325/Binding site: carbohydrate (Aen) (covalent) #status experimental

Query Match 77 %; Score 1023; DB 1; Length 458;  
Best Local Similarity 99.5%; Pred. No. 2, 6e-67;  
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MNRGVPFRHLVLTQLALPAATQGNKVVLGKKGDVETLTCTASQKKSIOFHMKNNOIK	60
DB	1	MNRGVPFRHLVLTQLALPAATQGNKVVLGKKGDVETLTCTASQKKSIOFHMKNNOIK	60
QY	61	ILGNQSPFLTKGSPSKINDRADSRRLMDQGNFPLIKNLKIEPSDYICVEVDQKEVOL	120
DB	61	ILGNQSPFLTKGSPSKINDRADSRRLMDQGNFPLIKNLKIEPSDYICVEVDQKEVOL	120
QY	121	LVFGLTRANSPDTHLQGSILTLTSPGSSPSVQCRSPGKNIQGGKTLVSQLELDPSG	180
DB	121	LVFGLTRANSPDTHLQGSILTLTSPGSSPSVQCRSPGKNIQGGKTLVSQLELDPSG	180
QY	181	TWTCITVLQNGQKVEFKIDIV	200
DB	181	TWTCITVLQNGQKVEFKIDIV	200

RESULT 2  
RMCTZ4  
T-cell surface glycoprotein CD4 - chimpanzee  
N/Alternate names: T-cell surface antigen T4/Les 3  
C/Species: Pan troglodytes (chimpanzee)  
C/Date: 30-Sep-1993 #sequence, revision 30-Sep-1993 #text, change 16-Jul-1999  
C/Accession: B32722; A46534  
R/Camerini, D.; Seed, B.  
Cell 60, 747-754, 1990  
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the v  
A/Reference number: A32722; MUID:90182664; PMID:2107024  
A/Accession: B32722  
A/Molecule type: mRNA  
A/Residues: 1-432 <CAM>  
A/Cross-references: GB:M31135



R:Pomgaard, A.; Hirsch, V.M.; Johnson, P.R.  
Eur. J. Immunol. 22, 2973-2981, 1992  
A:Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor  
A:Reference number: A46534; MUID:93049640; PMID:14425921  
A:Accession: A46534  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 3-399 <FOM>  
A:Note: sequence extracted from NCBI backbone (NCBI:118332)  
C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
F:1-371/Domain: extracellular #status predicted <EXT>  
F:9-86/Domain: immunoglobulin homology <IM1>  
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>  
F:191-274/Domain: immunoglobulin homology <IM3>  
F:296-347/Domain: immunoglobulin homology <IM4>  
F:372-395/Domain: transmembrane #status predicted <TM>  
F:396-432/Domain: intracellular #status predicted <INT>  
F:16-84,130-159,303-345/Diulfide bonds: #status predicted  
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 66.4%; Score 875; DB 1; Length 432;  
Best Local Similarity 97.7%; Pred. No. 1.5e-56;  
Matches 170; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 KVLVGGKGDVVELTCTASOKKSIQPHMKNNOIKILNQSGFLTKGPKLNDRAISRSL 86  
DB 2 KVLVGGKGDVVELTCTASOKKSIQPHMKNNOIKILNQSGFLTKGPKLNDRAISRSL 61  
QY 87 WDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSDTHLLQGSLTLTLESP 146  
DB 62 WDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSDTHLLQGSLTLTLESP 121  
QY 147 PGSSPSVQCRSPRKNIQGGKTLVSQLELQDSGTWCTVLOQNKVPEKIDIV 200  
DB 122 PGSSPSVQCRSPRKNIQGGKTLVSQLELQDSGTWCTVLOQNKVPEKIDIV 175

RESULT 3  
R:MOT4  
N:cell surface glycoprotein CD4 - rhesus macaque  
C:Alternate names: T-cell surface antigen T4/Leu 3  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
C:Accession: G32722  
R:Cammerlin, D.; Seed, B.  
Cell 60, 747-754, 1990  
A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi  
A:Reference number: A32722; MUID:90182664; PMID:2107024  
A:Accession: G32722  
A:Molecule type: mRNA  
A:Residues: 1-432 <CAM>  
A:Cross-references: GB:M1134  
C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
F:1-371/Domain: extracellular #status predicted <EXT>  
F:9-86/Domain: immunoglobulin homology <IM1>  
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>  
F:180-293/Domain: immunoglobulin homology <IM3>  
F:286-347/Domain: immunoglobulin homology <IM4>  
F:372-395/Domain: transmembrane #status predicted <TM>  
F:396-432/Domain: intracellular #status predicted <INT>  
F:16-84,130-159,303-345/Diulfide bonds: #status predicted  
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.0%; Score 790; DB 1; Length 432;  
Best Local Similarity 87.4%; Pred. No. 2.5e-50;  
Matches 152; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 27 KVLVGGKGDVVELTCTASOKKSIQPHMKNNOIKILNQSGFLTKGPKLNDRAISRSL 86  
DB 2 KVLVGGKGDVVELTCTASOKKSIQPHMKNNOIKILNQSGFLTKGPKLNDRAISRSL 61  
QY 87 WDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSDTHLLQGSLTLTLESP 146  
DB 62 WDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSDTHLLQGSLTLTLESP 121  
QY 147 PGSSPSVQCRSPRKNIQGGKTLVSQLELQDSGTWCTVLOQNKVPEKIDIV 200  
DB 122 PGSSPSVQCRSPRKNIQGGKTLVSQLELQDSGTWCTVLOQNKVPEKIDIV 175

RESULT 4  
A46254  
CD4 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jan-2000  
C:Accession: A46254  
R:Hague, B.F.; Sawadikhol, S.; Brown, T.J.; Lee, K.; Recker, D.P.; Kindt, T.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992  
A:Title: CD4 and its role in infection of rabbit cell lines by human immunodeficiency v  
A:Reference number: A46254; MUID:92390370; PMID:1518821  
A:Accession: A46254  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-459 <HNG>  
A:Cross-references: GB:M92840; NID:g164871; PIDN:AAA31198.1; PID:g164872  
A:Note: sequence extracted from NCBI backbone (NCBI:N112732, NCBI:P112733)  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
F:322-372/Domain: immunoglobulin homology <IM>

Query Match 45.3%; Score 596.5; DB 2; Length 459;  
Best Local Similarity 57.9%; Pred. No. 3.6e-36;  
Matches 124; Conservative 35; Mismatches 46; Indels 9; Gaps 3;

QY 1 MNRGVPRHLVLTALPAATQGNKVLGKGDVVELTCTASOKKSIQPHMKNNOIK 60  
DB 1 MNRGVPRHLVLTALPAATQGNKVLGKGDVVELTCTASOKKSIQPHMKNNOIK 60  
QY 61 ILGNQSGSSSSFWLKGNSPLSNVSEKKNMWDGSPFLVKDLAMDSDGYICEVEDQK 120  
DB 61 ILGNQSGSSSSFWLKGNSPLSNVSEKKNMWDGSPFLVKDLAMDSDGYICEVEDQK 120  
QY 117 EVOLLVGLTANSDTHLLQGSLTLTLESPGSSPSVQCRSPRKNIQGGKTLVSQLE 176  
DB 121 EVOLLVGLTANSDTHLLQGSLTLTLESPGSSPSVQCRSPRKNIQGGKTLVSQLE 180  
QY 177 QDSGTWCTVLOQNKVPEKIDIV----PRASA 205  
DB 181 QDSGTWCTVLOQNKVPEKIDIV----PRASA 214

RESULT 5  
S30193  
T-cell surface glycoprotein CD4 - dog  
C:Species: Canis lupus familiaris (dog)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jul-2000  
C:Accession: S30193  
R:Blidie, K.F.; Comer, G.E.; Mintz, D.H.; Alejandro, R.  
Biochim. Biophys. Acta 1172, 315-318, 1993  
A:Title: Primary structure of the canine CD4 antigen.  
A:Reference number: S30193; MUID:93192324; PMID:7916632  
A:Accession: S30193  
A:Molecule type: mRNA  
A:Status: preliminary  
A:Residues: 1-432 <ML>  
A:Cross-references: EMBL:X68565; NID:g288652; PIDN:CA837664.1; PID:g446737  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C:Keywords: glycoprotein  
F:202-311/Domain: immunoglobulin homology <IM>

Query Match 42.7%; Score 562; DB 2; Length 432;



```

QY      1  MNRGVPFRH-LLVLIQALALPAPATQGNKVLGKKGDVTELTCTASQOKKSIOFHMKNSNOI 59
      1  MCRALISLRLLILLQLQSLAVTQGTVLGKSGESALFPCSSQOKKITVFTKFSQDR 60
Db      60  KILOGNQG-SFLTKG--PSKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDOKE 116
QY      61  KILOGHKGVLIRGSGPSQF-DRFDSKKGAWEKGSFPLIKLIMEDSGQTYICELNRKE 119
Db      117  EVOLLVGLTRANSTHLLQGSLLTLLES-PPGSSPSVQCRSPGKXIOGKTLVSQOLE 175
QY      120  EVELMWKVFYFSPGTSLSLOGQSLTLTLDNSKNSVNPLETECKKXKGVSGSKVLSMSNLR 179
      176  LODSGTWTCTYLONOKK 192
QY      180  VQSDPFNNCTVTLDOCK 196
Db

RESULT 8
160082
CD4 receptor - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 23-Jul-1999
C/Accession: 160082
R/Zverev, V.V.; Sidorov, A.V.; Nedospasov, S.A.; Malinuhova, V.V.; Udaloova, I.A.; Andzha
Vopr. Virusol. 40, 100-102, 1995
A/Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene].
A/Reference number: 160082; MUID:95407135; PMID:7676667
A/Accession: 160082
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-71 <RES>
A/Cross-references: GB:579267; NID:g1086922; PIDN:AAB35273.1; PID:g1086923
C/Genetics:
A/Introns: 17/1
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

Query Match      27.1%; Score 357; DB 2; Length 71;
Best Local Similarity 96.6%; Pred. No. 1.2e-19;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  MNRGVPFRHLLVQLALLPAPATQGNKVLGKKGDVTELTCTASQOKKSIOFHMKNSNOI 60
      1  MNRGVPFRHLLVQLALLPAPATQGNKVLGKKGDVTELTCTASQOKKSIOFHMKNSNOI 60
Db      61  ILGNQGSFLTK 71
QY      61  ILGNQGSFLTK 71
      61  ILGNQGSFLTK 71
Db

RESULT 9
A39016
T-cell surface glycoprotein CD7 precursor - human
N/Alternate names: T-cell leukemia antigen
C/Species: Homo sapiens (man)
C/Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 05-Nov-1999
C/Accession: A39016; S03520
R/Schadberg, L.E.; Fleenor, D.E.; Kurtzberg, J.; Haynes, B.F.; Kaufman, R.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 603-607, 1991
A/Title: Isolation and characterization of the genomic human CD7 gene: structural simila
A/Reference number: A39016; MUID:91110576; PMID:1703303
A/Accession: A39016
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-240 <SCH>
A/Cross-references: GB:M37271; NID:g180163; PIDN:AAA51953.1; PID:g180164
R/Arnuffo, A.; Seed, B.
EMBO J. 6, 3313-3316, 1987
A/Title: Molecular cloning of two CD7 (T-cell leukemia antigen) cDNAs by a COS cell expr
A/Accession: S03520
A/Molecule type: mRNA
A/Residues: 1-240 <ARU>

```

```

A/Cross-references: EMBL:X06180; NID:g29819; PIDN:CAA29546.1; PID:g732757
C/Genetics:
A/Gene: GDB:CD7
A/Cross-references: GDB:119770; OMIM:186820
A/Map position: 17q25.2-17q25.3
A/Intons: 28/1
C/Keywords: glycoprotein; membrane protein; surface antigen; tandem repeat
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-240/Product: T-cell surface glycoprotein CD7 #status predicted <MNT>
F:145-180/Region: 9-residue repeats (X-P-P-X-A-S-A-L-P)

Query Match      25.2%; Score 332; DB 2; Length 240;
Best Local Similarity 36.4%; Pred. No. 3.5e-17;
Matches 102; Conservative 18; Mismatches 54; Indels 106; Gaps 10;

QY      6  PFRHLLVQLAL-LLPAPATQGNKVLGKK-----GDVTELTCTASQOKKSIOFHMKNS 56
      4  PFRHLLPRLALARGLPGLAAQEVQGSFHCCTVPVGAASVNTICSTSG----- 52
Db      57  NQIKILOGSFLTK-GPS-----KLNDRADSRSLMDQGNFPLIINK 98
QY      53  -----GLRGIVLRQLQEPQODIYYEDGVVPTTDRFRGRIDPSGS---QDNLITIMHR 103
      99  LKIEDSDTYICEVEDQKEEVQLVFGTLTANSTHLLQGSLLTLLESPPGSSPSVQCRSP 158
      104  LQSDTGTTC-----QATTEV----- 120
QY      159  RGNKIQGGKTLVSQLELQDSGTCTCTVLONQKKEFKDIVPAPALPAPPTGSALPDP 218
      121  ---NVVSGSLT-VLVTEEQSGQWHRCG-----DAPPASALPAPPTGSALPDP 164
      219  QPASALPDPAPASALPAPALAVISFLGLGIVACVLAART 258
      165  QPASALPDPAPASALPAPALAVISFLGLGIVACVLAARTQ 204
Db

RESULT 10
S21461
T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C/Accession: 147131; S21461
R/Gustafsson, K.; Germana, S.; Sundt, T.M.
J. Immunol. 151, 1365-1370, 1993
A/Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine
A/Reference number: 147131; MUID:93329116; PMID:8335933
A/Accession: 147131
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-99 <GU2>
A/Cross-references: EMBL:X65629; NID:g19228; PIDN:CAA46583.1; PID:g388232
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: glycoprotein; T-cell
F:3-81/Domain: immunoglobulin homology <1MM>

Query Match      23.2%; Score 305.5; DB 2; Length 99;
Best Local Similarity 60.2%; Pred. No. 1.1e-15;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY      32  KKGDTVTELTCTASQOKKSIOFHMKNSNOIKILOGNSFL-TKGPSKLNDRADSRSLMDQG 90
      1  KAGDLAEVLPCHSSQOKKNLPNWKNSNOTKILGSGHGFMTASTELTSLRDSKKNMWDHG 60
Db      91  NPLIINKLKIEDSDTYICEVEDQKEEVQLVFGTLN 128
QY      61  SFLIINKLLEVTDSGYICEVEDKRIEVLVRLTAS 98
      61  SFLIINKLLEVTDSGYICEVEDKRIEVLVRLTAS 98
Db

RESULT 11
S21462
T-cell surface glycoprotein CD4 (allele 2) - pig (fragment)
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000

```

[illegible]

M.; Kobayashi, M.; Asou, H.; Uyemura, K.  
 FEBS Lett. 289, 91-95, 1991  
 A>Title: Molecular cloning of cDNA encoding the rat neural cell adhesion molecule L1. T  
 A:Reference number: S17655; MUID:91372414; PMID:1894011  
 A:Accession: S36126  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1259 <MIU>  
 A:Cross-references: EMBL:X59149  
 A:Accession: S17655  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1178, 1183-1259 <MI2>  
 A:Cross-references: EMBL:X59149; NID:956740; PIDN:CAA41860.1; PID:956741  
 R:Prince, J.T.; Milona, N.; Stallcup, W.B.  
 J. Neurosci. 9, 1825-1834, 1989  
 A>Title: Characterization of a partial cDNA clone for the NILE glycoprotein and identifi  
 A:Reference number: A60917; MUID:88257627; PMID:2723751  
 A:Accession: A60917  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1159-1199, 'G', 1201-1235, 'K', 1237 <PRI>  
 A>Note: this paper appeared earlier, with printing errors, as reference A30326  
 R:Prince, J.T.; Milona, N.; Stallcup, W.B.  
 J. Neurosci. 9, 876-883, 1989  
 A>Title: Characterization of a partial cDNA clone for the NILE glycoprotein and identifi  
 A:Reference number: A30326; MUID:89177485; PMID:2466966  
 A:Contents: annotation  
 A>Note: this paper was reprinted as reference A60917 to correct the omission of several  
 C:Comment: This sequence of this surface-accessible glycoprotein differs at only two po  
 C:Superfamily: neural cell adhesion molecule L1, fibronectin type III repeat homology;  
 C:Keywords: cell adhesion; duplication; glycoprotein; membrane protein  
 F:531-592/Domains: immunoglobulin homology <IMV>  
 Query Match 10.1%; Score 133; DB 2; Length 1259;  
 Best Local Similarity 24.4%; Pred. No. 0.0086;  
 Matches 57; Conservative 31; Mismatches 96; Indels 50; Gaps 9;



A: Molecule type: mRNA  
 A: Residues: 1-739 <HES>  
 A: Cross-references: GB:M84488; NID:g207642; PIDN:AAA42332.1; PID:g207643  
 R:Williams, A.; Atkins, R.; Fries, J.; Gimbrone, M.A.; Cybulsky, M.I.; Collins, T.  
 Submitted to the EMBL Data Library, February 1992  
 A: Description: Nucleotide sequence of rat vascular cell adhesion molecule-1.  
 A: Reference number: S19872  
 A: Accession: S19872  
 A: Molecule type: mRNA  
 A: Residues: 1-2, 'G', 4-121, 'HL', 124-165, 'N', 167-738, 'G' <WIL>  
 A: Cross-references: EMBL:X63722; NID:g57471; PIDN:CAA45254.1; PID:g57472  
 R:Williams, A.J.; Atkins, R.C.; Fries, J.W.U.; Gimbrone Jr., M.A.; Cybulsky, M.I.; Collins, T.  
 Biochim. Biophys. Acta 1131, 214-216, 1992  
 A: Title: Nucleotide sequence of rat vascular cell adhesion molecule-1 cDNA.  
 A: Reference number: S23136; MUID:92305064; PMID:1377031  
 A: Accession: S23136  
 A: Status: preliminary  
 A: Molecule type: mRNA  
 A: Residues: 1-2, 'G', 4-165, 'N', 167-738, 'G' <WIL>  
 C: Comment: This protein interacts with the beta-1 integrin very late antigen 4 on leukocytes.  
 A: Gene: VCAM-1  
 C: Superfamily: immunoglobulin homology  
 C: Keywords: cell adhesion; transmembrane protein  
 F: 1-24/Domain: signal sequence #status predicted <SIG>  
 F: 25-739/Product: vascular cell adhesion molecule 1 #status predicted <VAS>  
 F: 259-293/Domain: immunoglobulin homology <IMM1>  
 F: 328-385/Domain: immunoglobulin homology <IMM2>  
 F: 526-581/Domain: immunoglobulin homology <IMM3>  
 F: 676-696/Domain: transmembrane #status predicted <TRA>  
 F: 697-715/Domain: intracellular #status predicted <INT>

Query Match 9.4%; Score 124; DB 2; Length 739;  
 Best Local Similarity 23.4%; Pred. No. 0.21;  
 Matches 43; Conservative 30; Mismatches 71; Indels 40; Gaps 5;

33 KGDVETCTCAQSKSIQPHMKNSNQIKLNGQSFITKPSKLNDRADRSIMDQNF 92  
 Db 238 EGAANVTTCASGEGPAEITFSKK-----LDNGVLQL-----SGNA 274  
 93 PLIINKLIEDSDTYICE---VEDQKEVQLV-----FGLTANSDFHLLQGOSLTLT 142  
 275 TLTLIARMEDSGIYVEGVNLVGRDTEVELIVQKRFPTVDISPGQVAAGVDSVLTL 334  
 143 LESPSSSPVQCSPPGKNIQ-----GKITSVQLBDSGTCTCYLONQKVER 195  
 Db 335 CAANGCSPFSWRTQTDSPLNGEVRDEGATLTLLSPVGEDEHSYLCVTCCRRLEK 394  
 196 KIDI 199  
 395 TIQV 398

RESULT 18  
 JC5519  
 50K glycoprotein precursor - chicken  
 C: Species: Gallus gallus (chicken)  
 C: Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 17-Nov-2000  
 C: Accession: JC5519  
 R: Hancock, K.A.; Gooley, A.A.; Jeffrey, P.L.  
 Mol. Brain Res. 44, 273-285, 1997  
 A: Title: ANGP50, a predominantly axonally expressed glycoprotein, is a member of the Ig1  
 A: Reference number: JC5519; MUID:97225899; PMID:9073169  
 A: Accession: JC5519  
 A: Molecule type: mRNA  
 A: Residues: 1-338 <HAN>  
 A: Experimental source: brain  
 C: Comment: This protein belongs to the IGLON's subfamily of cell adhesion molecules.  
 C: Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-terminal  
 C: Keyword: glycoprotein  
 F: 1-31/Domain: signal sequence #status predicted <SIG>  
 F: 32-338/Product: 50K glycoprotein #status predicted <MAT>  
 F: 46-113/Domain: immunoglobulin homology <IMM1>

F: 146-199/Domain: immunoglobulin homology <IMM2>  
 F: 232-292/Domain: immunoglobulin homology <IMM3>  
 F: 40, 136, 148, 279, 300, 315/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Query Match 9.3%; Score 123; DB 2; Length 338;  
 Best Local Similarity 24.1%; Pred. No. 0.097;  
 Matches 59; Conservative 41; Mismatches 107; Indels 38; Gaps 11;

10 LLIIVLQALLPAA-----TQGNKVLGKKDVTCTASQSKSIQPHMKNSQIKI 61  
 Db 14 LVTLRLTCLDPGLPVRVDFRTGDTNITVRQGDITLILCFVDRSS-KVAMNRSGLIF 72  
 62 LQNGQSFITKPSKLNDRAD-SRRSLMDQNEPLIINKLIEDSDTYICEVEDQ-----KE 116  
 73 AG-----EDKWSLDPRVELEKRSPL--YSLRIQKVDVDEGSITCSVGTQHHPKTS 122  
 117 EVQLVFG---LTNSDFHLLQGOSLTLTLESPSSSPVQCR--SPRKNIQGKT-L 169  
 Db 123 QVYLIQVPRPKISNISDITVNGSVTLVCMANGRPPIITWRHLPTGKEFEGEEVYL 182  
 170 SVSQLELDQSGTCTVQLQ---NOKVPEKIDIV--RASALPAPPTGSALPDTQAS 222  
 Db 183 ELIGITREQSGKYECKRANVASADVQVRYVTVNYPPTTESKSNBAATGQALLPCEAS 242  
 223 ALPDP 227  
 Db 243 AVETP 247

#### RESULT 19

A58532  
 glial cell membrane glycoprotein lig-1 precursor - mouse  
 C: Species: Mus musculus (house mouse)  
 C: Date: 11-Apr-1997 #sequence\_revision 11-Apr-1997 #text\_change 05-Nov-1999  
 C: Accession: A58532  
 R: Suzuki, Y.; Sato, N.; Tohyama, M.; Manaka, A.; Takagi, T.  
 J. Biol. Chem. 271, 22522-22527, 1996  
 A: Title: CDNA cloning of a novel membrane glycoprotein that is expressed specifically in  
 A: Reference number: A58532; MUID:96394313; PMID:8798419  
 A: Accession: A58532  
 A: Status: preliminary; translated from GB/EMBL/DBJ  
 A: Molecule type: mRNA  
 A: Residues: 1-1091 <SUZ>

A: Cross-references: GB:D78872; NID:g1545806; PIDN:BA11416.1; PID:g1545807  
 C: Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-  
 F: 36-61/Domain: proteoglycan amino-terminal homology <PAR>  
 F: 71-94/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>  
 F: 95-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>  
 F: 118-141/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>  
 F: 142-165/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>  
 F: 166-189/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>  
 F: 191-213/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>  
 F: 214-237/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>  
 F: 238-261/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>  
 F: 262-285/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>  
 F: 286-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>  
 F: 310-333/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>  
 F: 334-357/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>  
 F: 358-381/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>  
 F: 385-408/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>  
 F: 409-432/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>  
 F: 440-485/Domain: proteoglycan carboxyl-terminal homology <PCH>

Query Match 9.1%; Score 120; DB 2; Length 1091;  
 Best Local Similarity 22.3%; Pred. No. 0.64;  
 Matches 60; Conservative 40; Mismatches 103; Indels 66; Gaps 10;

14 LQIALPAAITQNGKVVILGKGDVETCTASQSKSIQPHMKNSNQIKLNGQSFITKGP 73  
 Db 595 LTVNVLDSFTKIRHDAIRGTTARLECAATGPNPQIAQKQGG-----TDFP 643  
 74 SKLNDRADRSIMDQNEPLIINKLIEDSDTYICEVEDQKEVQLVGLTANSDFHLL 133

Db 644 A-----ARERRHVMVDDDFEITDVKIDMGVSVCTAONSAGSV-----SANATLTV 691  
Qy 134 LOGOSLTLTLES-----PPGSSPVOCR-----SPRGKNIQGGKITS----- 170  
Db 692 LETPSLAVPLEDRVYVTEVAFOCKATGSPPTRTWIKGGRPLSTERRHFTPGNOILLV 751  
Qy 171 VSOLELDSGTWTCTVLONOKKVEFKIDIVPRA--SALPAPPGSALPDPQTASALPDPRA 229  
Db 752 VQNMWDAGRYTCE--MNPPLGTE-----RAHSQSLILPTPGCKKGTIVGIF----- 798  
Qy 230 ASALPALAVTSFLIGLGLGVAQVLAIRTR 258  
Db 799 -----TLAVVCSIVLTSLVWVCIIYQTR 821

RESULT 20  
T30532  
neural cell adhesion molecule l1 homolog - Fugu rubripes  
C/Species: Fugu rubripes  
C/Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 01-Mar-2002  
C/Accession: T30532  
R:Riboldi Tunicliffe, G.R.; Platzter, M.; Nakamura, G.; Elgar, G.S.; Brenner, S.; Rosen  
submitted to the EMBL Data Library, September 1997  
A/Description: Analysis of the genomic loci of Fugu rubripes homologs of the human disease  
A/Reference number: Z20848  
A/Accession: T30532  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1277 <RIB>  
A/Cross-references: EMBL:AF026198; NID:g3098263; PID:g3098264; PIDN:AAC1580.1  
C/Genetics:  
A/Intons: 42/1; 47/1; 81/2; 149/1; 190/1; 247/1; 285/2; 347/1; 391/1; 440/1; 477/2; 531/2  
A/Note: L1-CAM  
C/Superfamily: neural cell adhesion molecule l1; fibronectin type III repeat homology; 1

Query Match 9.1%; Score 120; DB 2; Length 1277;  
Best Local Similarity 24.7%; Pred. No. 0.78;  
Matches 55; Conservative 34; Mismatches 92; Indels 42; Gaps 9;

Qy 34 GDTVELTCTASQKSIQFHWKNSNQIKILNQGSFLTQPSKLNDRADSRSLMDQGNFP 93  
Db 363 GEYVLCQADGIGSPRTTW-----TVNGVPLSATSLSPRRSLPESGS-- 405  
Qy 94 LIINKLIKEDSDTYICEVEDQKE-----VQLVFGITANSDFH-LLOGSFLT-- 141  
Db 406 LILMDVIFGDTAIVQCCASNGKITLANTNYYIELPQILITNGNTYTFVEGQKALLEC 465  
Qy 142 -TESP-----PGSSPVOCRSPRGKNIQGGKITSVSOLELDSGTWTCTVLONOKKVE 194  
Db 466 ETFSPPKRYTWESSSISLLADPRVNLITNG-GLEIANVSHDEGIYTCLVQGSNINSVN 524  
Qy 195 FKIDIVPRASALPAPPGSALPDPQTASALPDPRA 233  
Db 525 AEVEVLNRTVTL-SPPQALRLQPKGTALFTCLVYTDPRLSPL 566

RESULT 21  
T13924  
sdc protein - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C/Accession: T13924  
R:Nguyen, D.N.; Liu, Y.; Litcky, M.L.; Reinke, R.  
submitted to the EMBL Data Library, February 1997  
A/Description: Sidekick, a member of the immunoglobulin superfamily, is required for pat  
A/Reference number: Z17809  
A/Accession: T13924  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-2222 <NGU>  
A/Cross-references: EMBL:U88578; NID:g4099554; PID:g4099555; PIDN:AAD09632.1  
C/Genetics:

A/Gene: sdc  
A/Cross-references: FlyBase:FBgn0021764

Query Match 9.1%; Score 120; DB 2; Length 2222;  
Best Local Similarity 24.0%; Pred. No. 1.5;  
Matches 54; Conservative 30; Mismatches 85; Indels 56; Gaps 11;

Qy 26 NKVYLGGKGDVELTCTASQKSIQFHWKNSNQIKILNQGSFLTQPSKLNDRADSRSLMDQGNFP 85  
Db 463 NVTALDGKADATISCRANGSPNPNTWYNETQLVDI-----SSIVQ 503  
Qy 86 LMDQGNPPLIINKLIKEDSDTYICEVEDQKEVQLVFGIT--ANSDFHLOGQ 137  
Db 504 ILESQD--LISNIRSVADAPLYICVRANVAGSVAAVLSVLTQIIQPPVDITVLIG- 560  
Qy 138 SLTLTLESPPGSSPV-----QCRSPRGKNIQGGKITSVSOLELDSGTWTCTVLONOKKVE 183  
Db 561 -LTATLTQCKVSSDPSVPYNIIDWREGQSTPISORIGVQADQLEIQAVRASDVGSYA 619  
Qy 184 CTVAQ-----NOKKVEFKIDIVPRASALPAPPGSALPDPQTAS 222  
Db 620 CVVTSPPGNETRAA-RLSVI-----ELPFPSPNVKVERLPEQNS 659

RESULT 22  
T08678  
hypothetical protein DKFZp56411922.1 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 23-Jul-1999  
C/Accession: T08678  
R:Mambur, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, May 1999  
A/Reference number: Z16469  
A/Accession: T08678  
A/Molecule type: mRNA  
A/Residues: 1-584 <WMB>  
A/Cross-references: EMBL:AL049946  
A/Experimental source: fetal brain; clone DKFZp56411922  
C/Genetics:  
A/Note: DKFZp56411922.1

Query Match 9.0%; Score 118.5; DB 2; Length 584;  
Best Local Similarity 21.2%; Pred. No. 0.4;  
Matches 59; Conservative 38; Mismatches 90; Indels 91; Gaps 12;

Qy 32 KKGDTVELTCTASQKSIQFHWKNSNQIKILNQGSFLT-----KGPSKLNDRADSRSLMDQGNFP 79  
Db 165 QRSDSGVYTCVLRNSAGEDRKTVI--VNVQPPKINGNPNPITVREIAGGSRKLTIDCK 223  
Qy 80 AD---SRSLMD-----QGNPPLIINKLIKEDSDTYICEVEDQKEV 118  
Db 224 AEGIPTRVLMAPPEGVLPAPYGNRTVHNGSLDIRSLKSDSVQLVCMANEGGEA 283  
Qy 119 QLT-----VFGITANSDFHLOGQSLTLTLESPPGSSPVOCRSPRGKNIQGG 166  
Db 284 RLIIQLTLVPEMEKPIRHDRISEKITMAAGHTISLNSAAGTIPPSLVWVLPNSTDIOG 343  
Qy 167 KTLT-----VSOLELDSGTWTCTVLON-----OKKVEFKIDIVPRAS 204  
Db 344 QQLQRFYHKADGMLHISGLSSVDAGAVRC-VARNAGHTRLVSLKVLGKPEASKQYHNL 402  
Qy 205 -----ALPAPPGSALPDPQTASALPDPRA 220  
Db 403 VSIINGETLKLPTCPGAGGGRPSWTLPLNGMHLEGPQT 440

RESULT 23  
S01998  
contactin precursor - chicken  
N/Alternate names: 130K glycoprotein  
C/Species: Gallus gallus (chicken)  
C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 17-Nov-2000  
C/Accession: S01998; J00094







```
Db      1465 IGIENKAEADYVEIDQAGLRGSCNVYLEKRPILNMKKIEALAG-EPCVVKVPFOI 1523
      211 TGSALDPDQ 219
      1524 KGTIRGDPK 1532

RESULT 28
503199
oploid-binding protein OPCAM precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 17-Mar-2000
C/Accession: S03199
R/Schrofield, P.R.; McFarland, K.C.; Haylick, J.S.; Wilcox, J.N.; Cho, T.M.; Roy, S.; Le
EMBO J. 8, 489-495, 1989
A/Title: Molecular characterization of a new immunoglobulin superfamily protein with pot
A/Reference number: S03199; MUID:9251576; PMID:2721489
A/Accession: S03199
A/Molecule type: mRNA
A/Residues: 1345 <SCH>
A/Cross-references: EMBL:X12672; NID:9585; PIDN:CAA31192.1; PID:9586
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C/Keywords: transmembrane protein
F/1-27/Domain: signal sequence #status predicted <SIG>
F/28-345/Product: oploid-binding protein OPCAM #status predicted <MAT>

Query Match      8.9%; Score 117.5; DB 2; Length 345;
Best Local Similarity 22.4%; Pred. No. 0.25;
Matches 59; Conservative 43; Mismatches 100; Indels 61; Gaps 10;

QY      11 LVVLQALLPAAQGNKVVVG-----KKGDTELTCTASQKSIQFMKNSN 57
      14 LVVSLRLFLVPTGVPRSGDATPPAMDNVTVQGESATLRCTIDRVT-RVAMLNRS 72
      58 QIKILGN-----QGSFLTKGSKLNDRAISRSLWDQGNFLLINKLIEDSDTYICE 110
      73 TLVAGNDKMSIDPRVILVNTPTQ-----YSLMGNVVDYEGPTCS 116
QY      111 VE----DQKEEVQLLVFG---LTNSDTHLLQGQSITLTLESPPGSSPSVQCRSPRGKN 162
      117 VQTDNHEKTSRVHLIVQVPPQIMNISDVTVNEGSSVTLTCLAIGRPEPVTWRHLSVKE 176
QY      163 IGG---GKTLVSQLELDQSGTWTCTVLO---NOKVYEFKIDYPRASALPAPPTGS 213
      177 GQGVSSDEYLEISDIRDSGGEYECALNDVAAPDVKVKITVNPYIS--KAKNTGV 234
Db      214 ALPDPTASALPDPPASALPAA 236
      235 SVQGKGLISC-----EASAVPMA 252

RESULT 29
A41288
vascular cell adhesion molecule 1, long splice form precursor - human
N/Alternate names: VCAM-1
C/Species: Homo sapiens (man)
C/Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 31-Jan-2000
C/Accession: A41288; S11476; A39755; B39755; A61160; A43352; PH1379; A39554
R/Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Eddy, R.; Byers, M.; Shows,
Proc. Natl. Acad. Sci. U.S.A. 88, 7859-7863, 1991
A/Title: Gene structure, chromosomal location, and basis for alternative mRNA splicing C
A/Reference number: A41288; MUID:91352090; PMID:11715583
A/Accession: A41288
A/Molecule type: DNA
A/Residues: 1-739 <CYB>
A/Cross-references: GB:M73255; NID:9340195; PIDN:AAA61270.1; PID:9340196
R/Polte, T.; Newman, M.; Gopal, T.V.
Nucleic Acids Res. 18, 5901, 1990
A/Title: Full length vascular cell adhesion molecule 1 (VCAM-1).
A/Reference number: S11476; MUID:91016951; PMID:1699207
A/Accession: S11476
A/Status: translation not shown

A/Molecule type: mRNA
A/Residues: 1-739 <BOU>
A/Cross-references: EMBL:X53051; NID:937648; PIDN:CAA37218.1; PID:937649
R/Hession, C.; Tizard, R.; Vassallo, C.; Schiffer, S.B.; Goff, D.; Moy, P.; Chi-Rosso, C.
J. Biol. Chem. 266, 6682-6685, 1991
A/Title: Cloning of an alternate form of vascular cell adhesion molecule-1 (VCAM1).
A/Reference number: A39755; MUID:91201302; PMID:1107873
A/Accession: A39755
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 25-646, 648-739 <HES1>
A/Cross-references: GB:M6035
A/Note: the complete translation is not shown
A/Accession: B39755
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-739 <HES2>
A/Cross-references: GB:M6035; NID:9340193; PIDN:AAA61269.1; PID:9340194
A/Experimental source: cell type endothelial cell; tissue type umbilical vein; map 1p32
R/Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Davis, V.M.; Glimbrone Jr.,
Am. J. Pathol. 138, 815-820, 1992
A/Title: Rapid communication. Alternative splicing of human VCAM-1 in activated vascula
A/Reference number: A61160; MUID:91189297; PMID:1107234
A/Accession: A61160
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 25-401, 'T', 403-686 <CY2>
R/Iademaro, M.F.; McQuillan, J.O.; Rosen, G.D.; Dean, D.C.
J. Biol. Chem. 267, 16323-16329, 1992
A/Title: Characterization of the promoter for vascular cell adhesion molecule-1 (VCAM-1)
A/Reference number: A43352; MUID:92355594; PMID:1379595
A/Accession: A43352
A/Molecule type: DNA
A/Residues: 1-21 <IAD>
A/Cross-references: GB:M92431; NID:9340197
A/Note: sequence extracted from NCBI backbone (NCBIN:110680, NCBI:P.110681)
R/Oborn, L.; Vassallo, C.; Benjamin, C.D.
J. Exp. Med. 176, 99-107, 1992
A/Title: Activated endothelium binds lymphocytes through a novel binding site in the al
A/Reference number: PH1379; MUID:92308860; PMID:1377228
A/Accession: PH1379
A/Molecule type: protein
A/Residues: 25-181, 'G', 183-402 <OSB>
C/Comment: This adhesion molecule is induced on endothelial cells by inflammatory cytok
C/Genetics:
A/Gene: GDB:VCAM1
A/Cross-references: GDB:127922; OMIM:192225
A/Map position: 1p32-1p31
C/Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein
F/1-24/Domain: signal sequence #status predicted <SIG>
F/25-739/Product: vascular cell adhesion molecule 1, long splice form #status predicted
F/25-698/Domain: extracellular #status predicted <EXT>
F/699-720/Domain: transmembrane #status predicted <TM>
F/721-739/Domain: intracellular #status predicted <INT>
F/273,365,417,463,531,561/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      8.9%; Score 117; DB 2; Length 739;
Best Local Similarity 22.3%; Pred. No. 0.67;
Matches 47; Conservative 38; Mismatches 84; Indels 42; Gaps 7;

QY      8 RHLLVQLALPAA--GQNVVLGKSDVTELTCTASQKSIQFMKNSNQIKILGNQ 65
      211 RQAVKELQVYISKINVISVNPETKLBEGGSVMTCSSEGLPAPLFW----- 258
Db      66 GSFVTGPKSLNDRASRSRLWDQGNFLLINKLIEDSDTYICE---VEDQKEVQL 121
      259 -----SKLDNGNMLQHL---SGNATLTLLAMRMEDSGIYCEGVNLGKRRKEVELI 307
QY      122 V-----FGITANSSTHLLQ--GQSITLTLESPPGSSPSVQCRSPRGKNIOG-----GKT 168
      308 VQKPTVEISPEPRIAQIGDSVMLTCSVMGCESSFSWRTQIDSPLSGKVRSEGTNST 367
Db      169 LSVSQLELDQSGTWTCTVLOQNKVEFKIDI 199
```

Db 368 LTLSPVSEFNEHSESYLCTVTCGHKKLEKGIQV 398

## RESULT 30

T13669

neuromusculin - fruit fly (*Drosophila melanogaster*)C:Species: *Drosophila melanogaster*

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: T13669

R:Kanla, A.; Han, P.L.; Kim, Y.T.; Bellen, H.

Neuron 11, 673-687, 1993

A:Title: Neuromusculin, a *Drosophila* gene expressed in peripheral neuronal precursors at

A:Reference number: Z17697; MUID:94000831; PMID:8398154

A:Accession: T13669

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1011 &lt;KAN&gt;

C:Genetics:

A:Cross-references: EMBL:L23146; NID:G385073; PID:G385074; PIDN:AAA03750.1

A:Gene: nrm

A:Cross-references: FlyBase:FBgn0005629

Query Match 8.9%; Score 117; DB 2; Length 1011;

Best Local Similarity 25.0%; Pred. No. 0.97;

Matches 48; Conservative 38; Mismatches 76; Indels 30; Gaps 9;

10 LLLVQLALPAATQGNKRVLLGKGGTVELCTASQKSIQPH---WKNSTQI--KIL 62

26 LVLVCLALVDSSTQAVDTTISQSGSGVLPCEPDAKCKGLHSLMFKGDDRIAMLL 85

63 GNGSFLTKGPKSNDRADSRSLMDQGNFPLIKNKIETSDTYICEVEDQKEVOLLV 122

86 GD-----SNVTSVAKKEDERTVY-EQNPYRLVIXDKLTAEDITLCTD-----T 128

123 FGLTANSDTHLQGSLLTLESPPGSSPSVQCSRPGKNIOGKTLVSQLELDSGTW 182

129 FPIPETCDN-FNGRIRLRVLPV---TEVVIDAKDKIRKNSV--VGMQGRQSLKA 182

183 TCTVQLNQKVE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

183 TCTVNRTRPPE 194

58 QIKLGN-----QGSFLTGPESKLNDRADSRSLMDQGNFPLIKNKIETSDTYICE 110

73 TLYAGNDKSIDRVLIVNTPTQ-----YSTIMQVNDYDECPYTCS 116

111 VE-----DQKEVOLLVFG---LTANSPTHLLQGSLLTLESPPGSSPSVQCSRPGKN 162

117 VQTNHRTKTSVHLIVQVPPQIMMISDITNESSVTLCLALGRPEPTVTHLSVKE 176

163 IQG---GKTLVSQLELDSGTWCTVLO-----NOKVEFKIDIVPRASALPAPPGS 213

177 GQGVSEDEYLEISDIRDQSGEYECALNDVAAPDVAKVITVNPYPYS--KAKYGV 234

214 ALPDPQASALPDPAPASALPAA 236

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

235 SVGQKGLISC---EASAVPMA 252

QY	202	R	202
Db	313	R	313

### RESULT 33

neural adhesion protein F3 precursor - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 20-Jun-2000  
C:Accession: J04211  
R:Watanabe, K.; Shimazaki, K.; Hosoya, H.; Fukumauchi, F.; Takenawa, T.  
Gene 160, 245-248, 1995  
A:Title: Cloning of the cDNA encoding neural adhesion molecule F3 from bovine brain.  
A:Reference number: J04211, MUID:95369697, PMID:7642103  
A:Accession: J04211  
A:Molecule type: mRNA  
A:Residues: 1-1018 <MAT>  
A:Cross-references: DDBJ:D32135, NID:g1060860, PIDN:BAA06861.; PID:g1060861  
A:Experimental source: brain  
C:Superfamily: contactin, fibronectin type III repeat homology; immunoglobulin homology  
C:Keywords: brain; glycoprotein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-1018/Product: neural adhesion protein F3 #status predicted <MAT>  
F:26-312/Domain: immunoglobulin homology <IMM>  
F:326-1018/Region: hydrophobic #status predicted  
F:358-457/Region: glycosylation site: N-linked  
F:208,358,338,457,473,494,521,591,630,933/Binding site: carbohydrate (Aen) (covalent) #

RESULT 34

secretory component precursor - rabbit  
N:Alternate names: poly-Ig receptor; polymeric immunoglobulin receptor  
N:Contains: free secretory component, transmembrane secretory component  
C:Species: *Oryctolagus cuniculus* (domestic rabbit)  
C:Date: 15-Nov-1994 #sequence\_revision 15-Nov-1994 #text\_change 16-Jul-1999  
A:Accession: A02111, A28077  
R:Mostov, K.E.; Friedlander, M.; Blobel, G.  
N:ature 308, 37-43, 1994  
A:Title: The receptor for trans epithelial transport of IgA and IgM contains multiple im  
A:Reference number: A02111, MUID:84142246, PMID:3322002  
A:Accession: A02111  
A:Molecule type: mRNA  
A:Residues: 1-773 <MOS>  
A:Cross-references: GB:X00412; GB:K01291; NID:g1595; PIDN:CA025118.1; PID:g1596  
A:Note: the authors translated the codon ACC for residue 54 as Asn  
R:Frutiger, S.; Hughes, G.J.; Hanly, W.C.; Jaton, J.C.  
R:BIOL. Chem. 265, 8120-8125, 1988  
A:Title: Rabbit secretory components of different allotypes vary in their carbohydrate

A;Reference number: A28077; MUID:88228032; PMID:31313339

A: Molecule type: protein  
A: Residues: 87-114,410-424 <FRU>  
C: Comment: This receptor binds polymeric IgA and IgM at the basolateral surface of epithelial cells.  
C: Comment: The five domains to separate the extracellular portion, also known as the secretory process, cleavage occurs to exhibit homology with immunoglobulin V regions. The similar  
C: Comment: Alternative splicing in the extracellular domain leads to high or low molecular weight isoforms.  
C: Superfamily: secretory component; immunoglobulin homology  
C: Keywords: alternative splicing; duplication; glycoprotein; immunoglobulin receptor; protein  
F: 1-18/Domain: signal sequence #status predicted <SIG>  
F: 19-773/Product: transmembrane secretory component #status predicted <MATM>  
F: 19-575/Product: free secretory component #status predicted <MATF>  
F: 30-647/Domain: extracellular #status predicted <EXT>  
F: 39-117/Domain: immunoglobulin homology <IM1>  
F: 148-227/Domain: immunoglobulin homology <IM2>  
F: 253-326/Domain: immunoglobulin homology <IM3>  
F: 362-440/Domain: immunoglobulin homology <IM4>  
F: 471-540/Domain: immunoglobulin homology <IM5>  
F: 648-670/Domain: transmembrane #status predicted <TMM>  
F: 671-773/Domain: intracellular #status predicted <INT>  
F: 46-115, 155-225, 260-324, 369-438, 478-538/Diulfide bonds: #status predicted  
F: 108/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental  
F: 418/Binding site: carboxylate (Asn) (covalent) #status experimental

Query Match	8.5%	Score 112.5;	DB 1;	length 773;
Best Local Similarity	23.4%	Pred. No. 1.5;		

### RESULT 35

Ig kappa chain V-J region (733-4) - human (Fragment)  
C:Species: Homo sapiens (man)  
C:Date: 27-Jan-1995 #sequence\_revision 01-Sep-1995 #text\_change 21-Jan-2000  
C:Accession: S46374; S38651  
R:Bensimon, C.; Chastagner, P.; Zouali, M.  
EMBO J. 13, 2951-2962, 1994  
A:Title: Human lupus anti-DNA autoantibodies undergo essentially primary V(ch) gene re  
A:Reference number: S46369; MUID:94313975; PMID:8039491  
A:Accession: S46374  
A:Molecule type: mRNA  
A:Residues: 1-120 <BEN>  
A:Cross-references: EMBL:Z27175; NID:G415965; PIDD:CAA01699.1; PIDD:G415966  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
C:KeyWords/Domain: immunoglobulin homology <IMM>



## RESULT 37

S33850  
fibronectin-binding protein - Streptococcus dysgalactiae

C:Species: Streptococcus dysgalactiae

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 15-Oct-1999

C:Accession: S33850, S33632

R:Lindgren, P.E.; McGavin, M.J.; Signaes, C.; Guss, B.; Gurnaidappa, S.; Hoceek, M.; Li

Eur. J. Biochem. 214, 819-827, 1993

A:Title: Two different genes coding for fibronectin-binding proteins from Streptococcus

A:Reference number: S33850; MUID:93307299; PMID:8316991

A:Accession: S33850

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1091 &lt;LIN&gt;

A:Cross-references: EMBL:Z22150; NID:9288966; PIDN:CAA80121.1; PID:9288969

Query Match 8.4%; Score 111; DB 2; Length 1091;

Best Local Similarity 23.0%; Pred. No. 2.9; Mismatches 66; Gaps 11;

Matches 64; Conservative 41; Mismatches 107; Indels 66; Gaps 11;

34 GDTVELCTASQKKSIGFHMKNNOIKIL---GNQSPFLTKGSKLNDRAISRSLMDQ 89

821 GQSEIRITEDSGSGMGQPGSGNETVEDTQSGEDIVLGGGQVIDFTEDSQPMMSG 880

90 GNPELLIKNLKIEDSDTYIC---EVEDQKEVQLVFGLTANSDTHL----- 134

881 NNSHTTEDSKPSQEDSVIIGGQGVDFTEDTQSGSGNSHTDGTVEEDSKPSQED 940

135 -----QGQSLTLTLESPGSSPSVQGRSP-----RGKNIQGGKTLVSQLELD--- 178

941 VIIGGQGVDFTEDTQSGAGQVSPITTEETHKPEILMGQSPIMWE--DTLPG 998

179 -SGTWCTVLQ--NQKVEFKIDIVPRASALPAP-PTGSALPPQRTASALPDPASALP 234

999 MSGSENVATVEEDTRPKLQFHFD---NEEPVATVPVSGTPPIAQSVSKVPHAKASALP 1055

235 -----AALAVTSFLGLGLGVAACVLAARR 258

1056 QTGDTNKLTFETFTTALTVI-----GAAGLGGKR 1085

## RESULT 38

I49443  
gene 284 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I49443

J. Immunol. 151, 5328-5337, 1993

A:Title: Cloning and characterization of the 284 gene encoding a molecule associated with

A:Reference number: I49443; MUID:94044757; PMID:8228228

A:Accession: I49443

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-398 &lt;RES&gt;

A:Cross-references: GB:I19057; NID:9309076; PIDN:AAA16353.1; PID:9309077

A:Genetics:

A:Gene: 284

Query Match 8.4%; Score 110.5; DB 2; Length 398;

Best Local Similarity 22.7%; Pred. No. 0.96; Mismatches 87; Indels 49; Gaps 9;

Matches 51; Conservative 38; Mismatches 87; Indels 49; Gaps 9;

18 LLPATOG-----NKVLGKKGDVVELCTASQKKSIGFHMKNNO-----IKILNQG 66

12 LLRAHQGQCPDSEEVVSGKPVQLRPSNIGTKOVSVQMKTEGSHRKIEIL----- 67

67 SFLTKGSKLNDRAISRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV-----QLL 121

68 NMYNDGFSWNSVPSFDIYG--FDYDGFALSTKSATLQDSGHVLEITVTGGVCKNQL 126

122 VFGLTANSDTHLQGGSLTLTLESPGSSPSVQ-----CRSPRGKNI----- 163

Db 127 IL-----DHY---ETPLKAKQKRWINGTQQLPLSLCVLTADDNVSYAFWRGSLTISN 176

Qy 164 QGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVPRASALPA 208

Db 177 QRNSTHENVQIDASSLHTYTCNVSRASWANTHLNTHGCGSVPS 221

## RESULT 39

TDFFLK

protein-tyrosine-phosphatase (EC 3.1.3.48) DLR precursor - fruit fly (Drosophila melan-

N:Alternate names: leukocyte antigen-related protein

C:Species: Drosophila melanogaster

C:Date: 14-Dec-1990 #sequence\_revision 02-May-1994 #text\_change 22-Jun-1999

C:Accession: A36182

R:Streuli, M.; Krueger, N.X.; Tsai, A.Y.M.; Saito, H.

Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989

A:Title: A family of receptor-linked protein tyrosine phosphatases in humans and Drosophi

A:Reference number: A36182; MUID:90046860; PMID:2554325

A:Accession: A36182

A:Molecule type: mRNA

A:Residues: 1-2029 &lt;STR&gt;

A:Cross-references: GB:M27700; NID:9157811; PIDN:AAA28668.1; PID:9157812

C:Genetics:

A:Gene: FlyBase:lar

A:Cross-references: FlyBase:FBgn0000464

C:Superfamily: Leukocyte antigen-related protein; fibronectin type III repeat homology;

ogy

C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane

F:1-32/Domain: signal sequence #status predicted &lt;SIG&gt;

F:33-2029/Product: leukocyte antigen-related protein #status predicted &lt;MAT&gt;

F:33-1377/Domain: extracellular #status predicted &lt;EXT&gt;

F:50-113/Domain: immunoglobulin homology &lt;IM1&gt;

F:154-211/Domain: immunoglobulin homology &lt;IM2&gt;

F:249-303/Domain: immunoglobulin homology &lt;IM3&gt;

F:321-401/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:416-502/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:514-599/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:610-699/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:708-802/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:811-896/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:909-993/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:1006-1091/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:1101-1198/Domain: fibronectin type III repeat homology &lt;FN3&gt;

F:1378-1402/Domain: transmembrane #status predicted &lt;TM&gt;

F:1403-2029/Domain: intracellular #status predicted &lt;INT&gt;

F:1417-2029/Domain: leukocyte common antigen cytosolic domain homology &lt;LAC&gt;

F:1487-1718/Domain: protein-tyrosine-phosphatase homology &lt;PTP1&gt;

F:1786-2009/Domain: protein-tyrosine-phosphatase homology &lt;PTP2&gt;

F:57-111,161-209,256-301/Dissulfide bonds: #status predicted

F:176/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1670/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1676/Binding site: substrate phosphate (Arg) #status predicted

F:1961/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1967/Binding site: substrate phosphate (Arg) #status predicted

Query Match 8.4%; Score 110.5; DB 1; Length 2029;

Best Local Similarity 23.0%; Pred. No. 6.6; Mismatches 50; Conservative 37; Mismatches 75; Indels 55; Gaps 11;

20 PATOGNKKVGLGKKGDVVELCTASQKKSIGFHMKNNOIKILNGGSLTKGSKLNDRA 79

140 PVTQPGRGVLEVHTVMTCAINPPTNIY-IKQTKV-----DM 182

80 ADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHL-LOGQS 138

183 SNRYSLKD--GR-LQIENSREDDQGYECVAENS-----MGTEHSAKTNLYVAVR 231

139 LITLESPP-----GSSPSVQGRS-----PRGNIGGKTL-----VSQ 174

232 VPETFRPPTTISEVMGLGSLNLSTAVSGPMPHVAKMGSEDLTENEMPGRNVLQ 291

175 ELQDSGTWCTVLONQKVEFKIDIVPRASALPAPT 211

Db 292 NIGESANTCTAAGTLCGID-SVSVV-KVQSLPRAFT 326

## RESULT 40

S34338

biliary glycoprotein P - mouse

N:Alternate names: mouse hepatitis virus (MHV) receptor glycoprotein

C:Species: Mus musculus (house mouse)

C&gt;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999

C/Accession: S34338, J01510, A41093

R: Huang, D. C.; Huang, X. F.; Novel, M.; Novel, G.

Submitted to the EMBL Data Library, July 1992

A:Description: A Clp-family gene present on the lactose-protease plasmid of lactococcus

A:Reference number: S34338

A:Accession: S34338

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-521 &lt;HUN&gt;

A:Cross-references: EMBL:X67281; NID:g312585; PIDN:CAA47698.1; PID:g312586

R:McNally, K.; Rosenberg, M.; Nedeliec, P.; Turbide, C.; Beauchemin, N.

Gene 127, 173-183, 1993

A:Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopro

A:Reference number: J01505; MUID:93273228; PMID:8500759

A:Accession: J01510

A:Molecule type: mRNA

A:Residues: 1-81, 'Q', 83-141, 'P', 143-521 &lt;MCC&gt;

A:Cross-references: GB:X67281

R:Williams, R. K.; Jiang, G. S.; Holmes, K. V.

Proc. Natl. Acad. Sci. U.S.A. 88, 5533-5536, 1991

A:Title: Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen

A:Reference number: A41093; MUID:91286498; PMID:1648219

A:Accession: A41093

A:Status: preliminary

A:Molecule type: protein

A:Residues: 35-59 &lt;WIL&gt;

C:Comment: This protein is expressed at the cell surface and plays a determinant role in

C:Genetics:

A:Gene: Bgpf

C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

C:Keywords: glycoprotein; receptor

F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology &lt;CEAN&gt;

F:160-219/Domain: immunoglobulin homology &lt;IMM1&gt;

F:254-303/Domain: immunoglobulin homology &lt;IMM2&gt;

F:339-386/Domain: immunoglobulin homology &lt;IMM3&gt;

F:87,104,148,199,206,210,226,258,290,294,304,333,375/Binding site: carbohydrate (Asn) (C

Query Match 8.4%; Score 110; DB 2; Length 521;

Best Local Similarity 23.7%; Pred. No. 1.4;

Matches 45; Conservative 34; Mismatches 69; Indels 42; Gaps 10;

QY 13 VLGLALPAAATQGNKVLGKKGVLELTCTA-SQKSIQFHKNSNOIKILGNQGSFLTK 71

Db 140 VHQLLKKNITSNNNSNPV-EGDDSVSLTCDSDYDPDNTYLSRN-----GESLSE 189

QY 72 GPKSLNDRADRRSLMOGNEPPLIIKNLKIDSDTYICEV-----DQKEVOL-LVFG- 124

Db 190 GD-----RLKLSGNRTLTLLANTNDTPYCEETNPVSVNRSDPFSNITLYGP 239

QY 125 ---LTANSDFHLQGSLLTLTSPGSSPSVQC-----RSPRKNIOGKTLISVSOLEL 176

Db 240 DTPILSPEDYILHFGSNLNLSCA--ASNPRAYFWLINEPHASS-----QELFIPMITT 293

QY 177 QDSGTWCTTV 186

Db 294 NNSGTYTCLV 303

## RESULT 41

JN0581

vascular cell adhesion molecule-1 long splice form precursor - mouse

C:Species: Mus musculus (house mouse)

C&gt;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 05-Nov-1999

C/Accession: JN0581; J06074; A40275; A48919

R:Araki, M.; Araki, K.; Vassalli, P.

Gene 126, 261-264, 1993

A:Title: Cloning and sequencing of mouse VCAM-1 cDNA.

A:Reference number: JN0581; MUID:93246254; PMID:7683304

A:Accession: JN0581

A:Molecule type: mRNA

A:Residues: 1-739 &lt;ARA&gt;

A:Cross-references: EMBL:X67783; NID:g298116; PIDN:CAA47989.1; PID:g298117

R:Hession, C.; Moy, P.; Tizard, R.; Christofm, P.; Williams, C.; Wyk, M.; Burki, L.; M

Biochem. Biophys. Res. Commun. 183, 163-169, 1992

A:Title: Cloning of murine and rat vascular cell adhesion molecule-1.

A:Reference number: J06074; MUID:92181437; PMID:1371918

A:Accession: J06074

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-739 &lt;HES&gt;

A:Cross-references: GB:M84487; NID:g202345; PIDN:AAA40545.1; PID:g202346

R:Myake, K.; Medina, K.; Ishihara, K.; Kimoto, M.; Auerbach, R.; Kinrade, P. W.

J. Cell Biol. 114, 557-565, 1991

A:Title: A VCAM-like adhesion molecule on murine bone marrow stromal cells mediates bin

A:Reference number: A40275; MUID:91317872; PMID:1713592

A:Accession: A40275

A:Molecule type: protein

A:Residues: 'XX', 27-32 &lt;MIY&gt;

R:Cybulsky, M. I.; Allan-McMamed, M.; Collins, T.

Genomics 18, 387-391, 1993

A:Title: Structure of the murine VCAM1 gene.

A:Reference number: A48919; MUID:94117008; PMID:7507076

A:Accession: A48919

A:Molecule type: DNA

A:Residues: 1-692, 'N', 694-739 &lt;CVB&gt;

A:Cross-references: GB:I22355; NID:g347981; PIDN:AAA16921.1; PID:g459893; GB:I22301

C:Comment: This protein is a transmembrane protein and interacts with the beta-1 integr

C:Superfamily: immunoglobulin homology

F:1-24/Domain: alternative splicing; cell adhesion; glycoprotein; transmembrane protein

F:25-739/Product: vascular cell adhesion protein-1 #status predicted &lt;MAT&gt;

F:239-283/Domain: immunoglobulin homology &lt;IMM1&gt;

F:338-385/Domain: immunoglobulin homology &lt;IMM2&gt;

F:527-581/Domain: immunoglobulin homology &lt;IMM3&gt;

F:659-720/Domain: transmembrane #status predicted &lt;TMN&gt;

F:225,273,424,531,561/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.4%; Score 110; DB 2; Length 739;

Best Local Similarity 21.3%; Pred. No. 2.2;

Matches 40; Conservative 34; Mismatches 68; Indels 46; Gaps 6;

QY 32 KKGPVLELTCTASQKSIQFHW--KNSNOI-KILGNQGSFLTKGPKSLNDRADRSRLMD 88

Db 237 QEGAVMTCTSSSEGLPAPBLFWGRKLDNEVLQL----- 270

QY 89 QGNFPLIIKNLKIDSDTYICE-----VEDQKEVOLIV-----FGLTANSDFHLQGS 138

Db 271 SGNAITLILMRNEDSVVYCEGVNLIGRKAEVELVQKPFIVDISPSQVAAQVGD 330

QY 139 LTLTLESPGSSPSVQCSPRGNIOG-----GKTLISVQLELQDSGTWCTTVLONOK 191

Db 331 VVLTCAIIGDSDPSFWSRTOTDPSPLNGVNVNEGAKSTLVLSVGFEDSHSLCAVTCLOR 390

QY 192 KVEFKIDI 199

Db 391 TLEKRTQV 398

## RESULT 42

S04693

T-cell receptor delta chain (NTD4) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C&gt;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 23-Jul-1999

C/Accession: S04693

R:Yoshikai, Y.; Matsuzaki, G.; Takeda, Y.; Ohga, S.; Kishihara, K.; Yuki, H.; Nomoto,

Eur. J. Immunol. 18, 1039-1043, 1988

A>Title: Functional T cell receptor delta chain gene messages in athymic nude mice.  
 A/Reference number: S04693; MUID:88296650; PMID:2969817  
 A/Accession: S04693  
 A/Molecule type: mRNA  
 A/Residues: 1-279 <YOS>  
 A/Cross-references: EMBL:X12729; NID:954894; PIDN:CAA1222.1; PID:954895  
 A/Note: this sequence was determined from the differentiated gene  
 C/Superfamily: Immunoglobulin V region; Immunoglobulin homology  
 C/Keywords: T-cell receptor

Query Match 8.3%; Score 109.5; DB 2; Length 279;  
 Best Local Similarity 21.8%; Pred. No. 0.75;  
 Matches 50; Conservative 29; Mismatches 85; Indels 65; Gaps 8;

QY 65 QGSELTGPGSKLNDRAISRSLMDQGNPILIKNKIEDSYIC---EVEDQKEVOL 120  
 DB 67 QGSEFVK-HSKAN-----RTFHLVISVSLSDSATYACASGISISGCGSTDK 113  
 QY 121 LVFGLTANSPTHLLOGLTLTLTSSPGSSPSVQCRRPKNI----- 163  
 DB 114 LVFG-----TGQVIVPEKSPAPKPSVFIMK-NGTNVACLVDKDFYKVTISLR 162  
 QY 164 -----OGKTLVSQLELDPSGTWTCTVLONOKKVEKIDIVPASALPA 208  
 DB 163 SSKKIVFDPRAIVISPGKXSVAVLGGYGDNSVTCSVCHNSERVH-STDFFPYANSF-- 219  
 QY 209 PPTGSAIPDPQTSALPDPAPASALPALAVISFLGLGACVCLART 257  
 DB 220 --NNKKLPEDNDTOISEPCGPRVIVHTEKVMNMLTVGLRLFAKT 266

## RESULT 43

S04663  
 T-cell receptor gamma chain - human (fragment)  
 C/Species: Homo sapiens (man)  
 C/Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jan-2000  
 C/Accession: S04663  
 R/Okada, A.; Bank, I.; Rogozinski, L.; Takihara, Y.; Mak, T.W.; Chees, L.; Alt, F.W.  
 J. Exp. Med. 168, 1481-1486, 1988  
 A/Title: Structure of the gamma/delta T cell receptor of a human thymocyte clone.  
 A/Reference number: S04663; MUID:89010543; PMID:2844954  
 A/Accession: S04663  
 A/Molecule type: DNA  
 A/Residues: 1-304 <OKA>  
 A/Cross-references: EMBL:X15018  
 A/Note: the authors translated the codon AAC for residue 72 and AAT for residues 102 and  
 C/Genetics:  
 A/Intons: 15/1; 133/3  
 C/Superfamily: Immunoglobulin V region; Immunoglobulin homology  
 C/Keywords: T-cell receptor  
 F,158-223/Domain: immunoglobulin homology <IMM>

Query Match 8.3%; Score 109.5; DB 2; Length 304;  
 Best Local Similarity 23.0%; Pred. No. 0.83;  
 Matches 47; Conservative 29; Mismatches 61; Indels 67; Gaps 9;

QY 8 RHLLVVLQALLPAPATGKNKVLGKGGDT-----VLLTCTASCKSKISQCHM----- 53  
 DB 2 RMALLVLLAFLSPASQSSNLEGGTKSVTRPSSAETCDLTVINAFYIHWYLHORGKA 61  
 QY 54 -----KNSNOIKILG-NOGSFLTGPGSKLNDRAISRSLMDQGNPILIKNKIED 103  
 DB 62 PQRLLVYDVNSKSVLESGLSPGKRYTHTP-----RRMSV-----ILILRLNLEND 107  
 QY 104 SDTYICEVEDQKEVOLLVFGLTANSPTHLLOGLTLT---LESPPGSSPSVQCRRPRG 160  
 DB 108 SGVYVCATWDRQDK--KLFG-----SGTTLVVTDKQADADVSPPRTFLP----- 150  
 QY 161 KNIQGGKTLVSQLELDPSGTWTC 184  
 DB 151 -----STAFETKQAKAGTYLC 165

## RESULT 44

A41060  
 neural cell adhesion molecule L1 precursor - human  
 N/Alternate names: L1CAM  
 C/Species: Homo sapiens (man)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C/Accession: A41060; S18454; A35331; S21971; S21972; A60223; A31072; G02506  
 R/Hayvin, M.L.; Lemmon, V.  
 Genomics 11, 416-423, 1991  
 A/Title: Molecular structure and functional testing of human L1CAM: an interspecies com  
 A/Reference number: A41060; MUID:92120663; PMID:1769655  
 A/Accession: A41060  
 A/Molecule type: mRNA  
 A/Residues: 1-1257 <HLA>  
 A/Cross-references: GB:M64296; NID:9186053; PIDN:AAC14352.1; PID:93068548  
 R/Kobayashi, M.; Miura, M.; Asou, H.; Uyemura, K.  
 Biochim. Biophys. Acta 1090, 238-240, 1991  
 A/Title: Molecular cloning of cell adhesion molecule L1 from human nervous tissue: a com  
 A/Reference number: S18454; MUID:92031698; PMID:1932117  
 A/Accession: S18454  
 A/Molecule type: mRNA  
 A/Residues: 1-3, 'V', 5-215, 'T', 217-249, 'T', 251-275, 'SV', 278-356, 'E', 358-625, 'V', 627-1257  
 A/Cross-references: EMBL:X59847; NID:935009; PIDN:CAA42508.1; PID:935010  
 A/Note: the authors translated the codon GAA for residue 27 as Gly  
 R/Djabali, M.; Matrei, M.G.; Nguyen, C.; Roux, D.; Demengeot, J.; Denizot, F.; Moos, M.,  
 Genomics 7, 587-593, 1990  
 A/Title: The gene encoding L1, a neural adhesion molecule of the immunoglobulin family,  
 A/Reference number: A35331; MUID:90353957; PMID:2387855  
 A/Accession: A35331  
 A/Molecule type: DNA  
 A/Residues: 332-371 <DTA>  
 A/Cross-references: GB:M55271  
 R/Rosenthal, A.; Mackinnon, R.N.; Jones, D.S.C.  
 Nucleic Acids Res. 19, 5395-5401, 1991  
 A/Title: PCR walking from microdissection clone M54 identifies three exons from the hum  
 A/Reference number: S21971; MUID:92020233; PMID:1923824  
 A/Accession: S21971  
 A/Molecule type: DNA  
 A/Residues: 1082-1176 <ROS>  
 A/Cross-references: EMBL:X58775; NID:929642; PIDN:CAA1576.1; PID:929643  
 A/Accession: S21972  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 353-935, 'V', 937-1176 <RO2>  
 A/Cross-references: EMBL:X58776; NID:929644; PIDN:CAA37831.1; PID:94467833  
 R/Hatper, J.R.; Prince, J.T.; Healy, P.A.; Stuart, J.K.; Nauman, S.J.; Stallcup, W.B.  
 J. Neurochem. 56, 797-804, 1991  
 A/Title: Isolation and sequence of partial cDNA clones of human L1: homology of human a  
 A/Reference number: A60223; MUID:9132183; PMID:1993895  
 A/Accession: A60223  
 A/Status: not compared with conceptual translation  
 A/Molecule type: protein  
 A/Residues: 1030-1115, 'WLC', 1118-1176, 1181-1257 <HAR>  
 R/Wolfe, J.M.; Frank, R.; Mujo, K.; Spiro, R.C.; Reisfeld, R.A.; Rathjen, F.G.  
 J. Biol. Chem. 263, 11943-11947, 1988  
 A/Title: A human brain glycoprotein related to the mouse cell adhesion molecule L1.  
 A/Reference number: A31072; MUID:88298876; PMID:3136168  
 A/Accession: A31072  
 A/Molecule type: protein  
 A/Residues: 'Q', 21-36 <WOL>  
 R/Platzner, M.; Bauer, D.; Drescher, B.  
 submitted to the EMBL Data Library, March 1995  
 A/Reference number: H01368  
 A/Accession: G02506  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1257 <PLA>  
 A/Cross-references: EMBL:U52112; NID:91302657; PIDN:AAC51746.1; PID:91302658  
 C/Genetics:  
 A/Gene: GDB:L1CAM  
 A/Cross-references: GDB:120133; OMIM:303350; OMIM:308840  
 A/Map position: Xq28-Xq28  
 A/Intons: 26/1; 31/1; 66/2; 134/1; 175/1; 232/1; 269/2; 331/1; 375/1; 423/1; 460/2; 51



/2  
 C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; i  
 C:Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmembran  
 P:1-19/Domains: signal sequence #status predicted <SIG>  
 F:20-1257/Product: neural cell adhesion molecule L1 #status predicted <MAM>  
 F:257-314/Domains: immunoglobulin homology <IMM1>  
 F:533-593/Domains: immunoglobulin homology <IMM2>

Query March 8.3%; Score 109; DB 1; Length 1257;  
 Best Local Similarity 20.9%; Pred. No. 4.8;  
 Matches 48; Conservative 35; Mismatches 97; Indels 50; Gaps 9;

QY 8 RHLLLV-----LQALLPAA--ATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNQKIQI 61  
 DB 409 RHGULLAAAYIVVQLPKILTLADNQTMAVQSGFAYLLCKAFGAPVSVQMLDDEDTTV 468  
 QY 62 LGNGGSLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVL 121  
 DB 469 LQDERFF-----PYANGFLGIRDLQANDTGRYFCLANDQNNVTIM 509  
 QY 122 VFGLTANDTHLQ-----GSLTLTLES--PRSSSPVQCRSFRKNIQ----- 164  
 DB 510 A-NLKVKADATQITQGPSTRTKESGRVFTQASFPDLQPSITWRGD-GRDLQELGSD 567  
 QY 165 -----GKTLVSQLELDQSGTWTCTVLQNGKVEFKIDIVPRASALPAP 209  
 DB 568 KYFIEDGR-LVTHSDYSDQGNVSCVASTELDVAESRAQLVGSFGFVP 616

#### RESULT 45

protein-tyrosine kinase (EC 2.7.1.112) Flk-1 precursor, endothelial cell-specific recept  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Jun-1992 #sequence\_revision 19-Jun-1992 #text\_change 04-Feb-2000  
 C:Accession: A41228; A46065; I58365; S18832; S29991

R:Matthews, W.; Jordan, C.T.; Gavin, M.; Jenkins, N.A.; Copeland, N.G.; Lemischka, I.R.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 9026-9030, 1991  
 A:Title: A receptor tyrosine kinase cDNA isolated from a population of enriched primitiv  
 A:Reference number: A41228; MUID:92020984; PMID:1717995  
 A:Accession: A41228  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1367 <MAM>  
 A:Cross-references: GB:X59397; NID:G50976; PIDN:CAA42040.1; PID:G50977  
 R:Miller, B.; Witzmann-Voos, S.; Schmitt, H.; Martinez, R.; Moller, N.P.; Rieau, W.;  
 Cell 72, 835-846, 1993  
 A:Title: High affinity VEGF binding and developmental expression suggest Flk-1 as a maj  
 A:Reference number: A46065; MUID:93208880; PMID:7681362  
 A:Accession: A46065  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-74, 'T', '26-782, 'VL', '785-916, 'C', '918-1367 <MIL>  
 A:Cross-references: GB:X70842; NID:G57923; PIDN:CAA50192.1; PID:G57924  
 A:Note: submitted to the EMBL Data Library, January 1993  
 A:Note: sequence extracted from NCBI backbone (NCBI:128066)  
 R:Oelrichs, R.B.; Reid, H.H.; Bernard, O.; Ziemlecki, A.; Wils, A.F.  
 Oncogene 8, 11-18, 1993  
 A:Title: NYK/Flk-1: a putative receptor protein tyrosine kinase isolated from E10 embry  
 A:Reference number: 158365; MUID:93141255; PMID:8423988  
 A:Accession: 158365  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-578, 'D', '680-1340, 'RSPV', <OEL>  
 A:Cross-references: GB:S53103; NID:G264004; PIDN:AAB25043.1; PID:G264005  
 C:Genetics:  
 A:Gene: FLK-1; NYK  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo  
 C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; transmembrane  
 F:830-1165/Domains: protein kinase homology <KIN>  
 F:83-846/Region: protein kinase ATP-binding motif

Query March 8.3%; Score 109; DB 2; Length 1367;  
 Best Local Similarity 21.0%; Pred. No. 5.3;

Matches 45; Conservative 36; Mismatches 73; Indels 60; Gaps 9;  
 QY 9 HLLVLQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNQKIQI 68  
 DB 544 HVIGPEITVQPAQPTQ-----BSVSLCTADRTFENLTWK-----LGSQYTS 590  
 QY 69 LTKPSKLNDRADSRSLM-----DQGNFPLI--KNLKIEDSDTYICEVEDQKEE 117  
 DB 591 VHMESS-LTPVCKRDLMLKNGTMFSNSTDILIVAFQNASLDQGDYVCSADKTKK 649  
 QY 118 -----VQLVFGLTANDTHLQGSLLTLESPPGSSPVQCRSPKGN----- 162  
 DB 650 RHCLVKQILIERMAPMITGNLENQTTI-----GETIEVTC--PASGNPTPIITWKD 701  
 QY 163 -----IQGKTLVSQLELDQSGTWTCT 164  
 DB 702 NETIVEDSGIVLRDGNRLTIRVRKEDGGLYTC 735

#### RESULT 46

A31923  
 Amalgam protein precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 21-Jul-2000  
 C:Accession: A31923  
 R:Seeger, M.A.; Haefley, L.; Kaufman, T.C.  
 Cell 55, 589-600, 1988  
 A:Title: Characterization of amalgam: a member of the immunoglobulin superfamily from D  
 A:Reference number: A31923; MUID:89028670; PMID:3141062  
 A:Accession: A31923  
 A:Molecule type: DNA  
 A:Residues: 1-333 <SSE>  
 A:Cross-references: GB:M2361; NID:G156920; PIDN:AAA28367.1; PID:G156921  
 C:Genetics:  
 A:Gene: FlyBase:Ana  
 A:Cross-references: FlyBase:FBgn0000071

Query March 8.2%; Score 108.5; DB 2; Length 333;  
 Best Local Similarity 20.8%; Pred. No. 1.1;  
 Matches 49; Conservative 38; Mismatches 90; Indels 59; Gaps 10;

QY 10 LLLVLQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHM-----KNSNQ 58  
 DB 10 LIFCLALSLSVLSAPYISQSKVVASGVDFNCTVEVGLSVSMARPSBSPTNS 69  
 QY 59 I-----KILG---NQGSLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYIC 109  
 DB 70 VLISMRLILSLPDKRYVWVTEGPKT-----GSAITYFRIQNIIEVSDMGPEYC 117  
 QY 110 EV-----EDQKEEVL---VFGLTANDTHLQGSLLTLESPPGSSPV----- 153  
 DB 118 QVLVSATEKVKKLSIDIKTPPVIAENTPKSTLTVEQNLELTHANGFPKPTISWAREH 177  
 QY 154 QCRSPKGNQGGKTLVSQLELDQSGTWTCTVLQNGKVEFKIDIVPRASALPAP 199  
 DB 178 NAWNPAGHLLAEPTLIRSVHRMDRGYYC-IAQNEGQPDRLIVVEVFRQI 232

#### RESULT 47

S00682  
 IGE FC receptor alpha chain precursor - human

N:Alternate names: Fc-epsilon receptor  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 23-Jul-1999  
 C:Accession: S00682; B30154; S42209  
 R:Kochan, J.; Petrine, L.F.; Hakim, K.; Kline, J.P.  
 Nucleic Acids Res. 16, 3584, 1988  
 A:Title: Isolation of the gene coding for the alpha subunit of the human high affinity  
 A:Reference number: S00682; MUID:88233953; PMID:2967464  
 A:Accession: S00682  
 A:Molecule type: mRNA  
 A:Residues: 1-257 <KOC>  
 A:Cross-references: EMBL:X06948; NID:G31317; PIDN:CAA30025.1; PID:G31318

R.Shimizu, A.; Tepler, I.; Benfey, P.N.; Berenslein, E.H.; Siraganian, R.P.; Leder, P.  
Proc. Natl. Acad. Sci. U.S.A. 85, 1907-1911, 1988  
A>Title: Human and rat mast cell high-affinity immunoglobulin E receptors: characterized  
A/Reference number: A94191; MUID:88158102; PMID:2964640  
A/Accession: B30154  
A/Molecule type: mRNA  
A/Residues: 1-257 <SH1>  
A/Cross-references: GB:J03605; NID:g187449; PIDN:AAA36204.1; PID:g307164  
R.Yagi, S.; Yanagida, M.; Tanida, I.; Hasegawa, A.; Okumura, K.; Ka, C.  
Eur. J. Biochem. 220, 593-598, 1994  
A>Title: High-level expression of the truncated alpha chain of human high-affinity receptor product.  
A/Reference number: S42209; MUID:94170811; PMID:8125119  
A/Accession: S42209  
A/Molecule type: protein  
A/Residues: 26-197 <YAG>  
A/Experimental source: purified recombinant protein  
C/Genetics:  
A/Gene: GDB:FCER1A  
A/Cross-references: GDB:119902; OMIM:147140  
A/Map position: 1q23-1q23  
C/Superfamily: FC gamma receptor III; immunoglobulin homology  
C/Keywords: immunoglobulin receptor; transmembrane protein  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-257/Product: IGE FC receptor alpha chain #status predicted <MAT>  
F:44-95/Domain: immunoglobulin homology <IMM1>  
F:125-178/Domain: immunoglobulin homology <IMM2>

Query Match 8.2%; Score 108; DB 2; Length 257;  
Best Local Similarity 27.1%; Pred. No. 0.88;  
Matches 55; Conservative 22; Mismatches 64; Indels 62; Gaps 10;

OY 10 LLLVQLALP-----AATGKRVVGLK-----KGTVELTCTASQKSIQFHWKNSNQI 59  
10 LLLVQLALP-----AATGKRVVGLK-----KGTVELTCTASQKSIQFHWKNSNQI 51  
10 LLLVQLALP-----AATGKRVVGLK-----KGTVELTCTASQKSIQFHWKNSNQI 51  
10 LLLVQLALP-----AATGKRVVGLK-----KGTVELTCTASQKSIQFHWKNSNQI 51

OY 60 KIINGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117  
60 KIINGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117  
60 KIINGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117

OY 52 ----NGNPFREVSSTKPFHNG----SLSEETNSSLNINAKFEDSGYKCKQHQQVNESEP 103  
52 ----NGNPFREVSSTKPFHNG----SLSEETNSSLNINAKFEDSGYKCKQHQQVNESEP 103  
52 ----NGNPFREVSSTKPFHNG----SLSEETNSSLNINAKFEDSGYKCKQHQQVNESEP 103

OY 118 VOLLVFGILTANSDTHLQGSLLTLTLESPGSSPVOCRSPPRGNI-----QGGKTL-- 169  
118 VOLLVFGILTANSDTHLQGSLLTLTLESPGSSPVOCRSPPRGNI-----QGGKTL-- 169  
118 VOLLVFGILTANSDTHLQGSLLTLTLESPGSSPVOCRSPPRGNI-----QGGKTL-- 169

OY 104 VYLEVF-----SDMLLLQA--SAEYVMEGQF---LFLKCHGMNDVYKVIYVDGEALYK 154  
104 VYLEVF-----SDMLLLQA--SAEYVMEGQF---LFLKCHGMNDVYKVIYVDGEALYK 154  
104 VYLEVF-----SDMLLLQA--SAEYVMEGQF---LFLKCHGMNDVYKVIYVDGEALYK 154

OY 170 -----SVQLELQDSGTWTCT 185  
170 -----SVQLELQDSGTWTCT 185  
170 -----SVQLELQDSGTWTCT 185

OY 155 WYENHNISITNATVEDSGTYTCT 177  
155 WYENHNISITNATVEDSGTYTCT 177  
155 WYENHNISITNATVEDSGTYTCT 177

RESULT 48  
S40370  
IG kappa chain - human  
C/Species: Homo sapiens (man)  
C/Date: 06-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 21-Jan-2000  
C/Accession: S40370  
R.Klein, R.; Jaenichen, R.; Zachau, H.G.  
Eur. J. Immunol. 23, 3248-3271, 1993  
A>Title: Expressed human immunoglobulin chi genes and their hypermutation.  
A/Reference number: S40312; MUID:94080891; PMID:8258341  
A/Accession: S40370  
A/Status: preliminary; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-122 <KLE>  
A/Cross-references: EMBL:X72480; NID:g441428; PIDN:CAA51148.1; PID:g441429  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterocramer; immunoglobulin  
F:30-104/Domain: immunoglobulin homology <IMM>

Query Match 8.2%; Score 107.5; DB 2; Length 122;  
Best Local Similarity 35.5%; Pred. No. 0.39;  
Matches 38; Conservative 5; Mismatches 53; Indels 11; Gaps 3;

OY 11 LVLVQLALPAA-----TQGNKVVLGKGDVLELTCTASQKSIQFHWKNSNQIKILGN 64  
11 LVLVQLALPAA-----TQGNKVVLGKGDVLELTCTASQKSIQFHWKNSNQIKILGN 64  
11 LVLVQLALPAA-----TQGNKVVLGKGDVLELTCTASQKSIQFHWKNSNQIKILGN 64

OY 1 LGLLLMLLRGASCDIQLTQSSSLASVGDVTTTCRASQSIPTLHYQON----LGK 56  
1 LGLLLMLLRGASCDIQLTQSSSLASVGDVTTTCRASQSIPTLHYQON----LGK 56  
1 LGLLLMLLRGASCDIQLTQSSSLASVGDVTTTCRASQSIPTLHYQON----LGK 56

OY 65 QGSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICE 110  
65 QGSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICE 110  
65 QGSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICE 110

OY 57 APRLLIYASNLQSGVSRFSGSGSDTFTLTISGLQDPDFATYYCQ 103  
57 APRLLIYASNLQSGVSRFSGSGSDTFTLTISGLQDPDFATYYCQ 103  
57 APRLLIYASNLQSGVSRFSGSGSDTFTLTISGLQDPDFATYYCQ 103

RESULT 49  
IHHUNG  
neural cell adhesion molecule 1 GPI-anchored splice form precursor, muscle-specific - h  
N/Alternate names: CD56; NCAM-120  
C/Species: Homo sapiens (man)  
C/Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 28-Jan-2000  
A/Accession: S07784; A26883  
R.Barton, C.H.; Dickson, G.; Gower, H.J.; Rowett, L.H.; Putt, W.; Elsom, V.; Moore, S.E  
Development 104, 165-173, 1988  
A>Title: Complete sequence and in vitro expression of a tissue-specific phosphatidylinositol  
A/Reference number: S07784; MUID:89305258; PMID:3253057  
A/Accession: S07784  
A/Molecule type: mRNA  
A/Residues: 1-761 <BAR>  
A/Cross-references: EMBL:X16841; NID:g35005; PIDN:CAA34739.1; PID:g35006  
R.Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; Cox, Cell 50, 1119-1130, 1987  
A>Title: Human muscle neural cell adhesion molecule (N-CAM): identification of a muscle  
A/Reference number: A90895; MUID:87301755; PMID:2887295  
A/Accession: A26883  
A/Molecule type: mRNA  
A/Residues: 491-761 <DIC>  
A/Cross-references: GB:M17409; NID:g189097; PIDN:AAA59912.1; PID:g386979  
C/Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM moi  
C/Genetics:  
A/Gene: GDB:NCAM1; NCAM; CD56  
A/Cross-references: GDB:119448; OMIM:116930  
A/Map position: 11q22.2-11q22.3  
C/Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; imm  
C/Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, muscl  
F:34-98/Domain: immunoglobulin homology <IMM1>  
F:132-191/Domain: immunoglobulin homology <IMM2>  
F:152-156/Region: heparin binding #status predicted  
F:161-165/Region: heparin binding #status predicted  
F:228-289/Domain: immunoglobulin homology <IMM3>  
F:263-272/Region: NCAM binding #status predicted  
F:322-387/Domain: immunoglobulin homology <IMM4>  
F:419-481/Domain: immunoglobulin homology <IMM5>  
F:499-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>  
F:633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>  
F:41-96,139-189,235-287,328-385,426-479/Disulfide bonds: #status predicted  
F:222,315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.1%; Score 107; DB 1; Length 761;  
Best Local Similarity 19.3%; Pred. No. 3.7;  
Matches 44; Conservative 38; Mismatches 102; Indels 44; Gaps 7;

OY 3 RG-VPRHLLVQLALPAAATGKRVVLGKGDVLELTCTASQKSIQFHWKNSNQIKI 61  
3 RG-VPRHLLVQLALPAAATGKRVVLGKGDVLELTCTASQKSIQFHWKNSNQIKI 61  
3 RG-VPRHLLVQLALPAAATGKRVVLGKGDVLELTCTASQKSIQFHWKNSNQIKI 61

OY 196 RGEINFKDIOIVNVPPPTIQARQINATANIGQSVTLVCDAGPEPTMSW----- 247  
196 RGEINFKDIOIVNVPPPTIQARQINATANIGQSVTLVCDAGPEPTMSW----- 247  
196 RGEINFKDIOIVNVPPPTIQARQINATANIGQSVTLVCDAGPEPTMSW----- 247

OY 62 LQNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117  
62 LQNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117  
62 LQNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLIKEDSDTYICEVE--DQKEE 117

OY 248 -----TKDGEQIQEEDDEKTYIFSDSSQLTIKKVKNDAEYICIAENKAGQDAT 299  
248 -----TKDGEQIQEEDDEKTYIFSDSSQLTIKKVKNDAEYICIAENKAGQDAT 299  
248 -----TKDGEQIQEEDDEKTYIFSDSSQLTIKKVKNDAEYICIAENKAGQDAT 299

OY 118 VOLLVFG-----LTANSDTHLQGSLLTLTLESPGSSPVOCR-----SPRGNIQGG- 166  
118 VOLLVFG-----LTANSDTHLQGSLLTLTLESPGSSPVOCR-----SPRGNIQGG- 166  
118 VOLLVFG-----LTANSDTHLQGSLLTLTLESPGSSPVOCR-----SPRGNIQGG- 166

OY 300 IHLKVFAPKRTIVVENQTMLEBEQVTLTCEASGDPDPSTWTSTRNISSEKTLDDGM 359  
300 IHLKVFAPKRTIVVENQTMLEBEQVTLTCEASGDPDPSTWTSTRNISSEKTLDDGM 359  
300 IHLKVFAPKRTIVVENQTMLEBEQVTLTCEASGDPDPSTWTSTRNISSEKTLDDGM 359

OY 167 -----KTLVSQLELQDSGTWTCT-----TVLQNGKRVYEFKIDIVPR 202  
167 -----KTLVSQLELQDSGTWTCT-----TVLQNGKRVYEFKIDIVPR 202  
167 -----KTLVSQLELQDSGTWTCT-----TVLQNGKRVYEFKIDIVPR 202

OY 360 VVASHARVSVSLTLKSIQYTDAGYICTASNTIGQDSQSWYLEVQVAPK 407  
360 VVASHARVSVSLTLKSIQYTDAGYICTASNTIGQDSQSWYLEVQVAPK 407  
360 VVASHARVSVSLTLKSIQYTDAGYICTASNTIGQDSQSWYLEVQVAPK 407

```

RESULT 50
S18252
heparan sulfate proteoglycan - mouse
N/Alternate names: perlecan
C/Species: Mus musculus (house mouse)
C/Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 05-Nov-1999
C/Accession: S18252; A31917; S66460
R/Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A/Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogly
adhesion molecule.
A/Reference number: S18252; MUID:92078153; PMID:1744087
A/Accession: S18252
A/Molecule type: mRNA
A/Residues: 1-3707 <NOO>
A/Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R/Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hase
J. Biol. Chem. 263, 16379-16387, 1988
A/Title: Identification of cDNA clones encoding different domains of the basement membra
A/Reference number: A52680; MUID:89034110; PMID:2972708
A/Accession: A31917
A/Molecule type: mRNA
A/Residues: 940-1601 <NO2>
A/Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A/Accession: B31917
A/Molecule type: mRNA
A/Residues: 1870-2600 <NO3>
A/Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R/Schulze, B.; Mann, K.; Bartschke, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A/Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A/Reference number: S66460; MUID:95377282; PMID:7649154
A/Accession: S66460
A/Molecule type: protein
A/Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
C/Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C/Keywords: glycoprotein
F/199-224/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F/283-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F/325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F/368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F/764-811/Domain: laminin-type EGF-like homology <LEG>
F/1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F/1563-1610/Domain: laminin-type EGF-like homology <LEG8>
F/1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F/3163-3198/Domain: EGF homology <EGF>
F/3270-3423/Domain: laminin G repeat homology <LG2>
F/3464-3492/Domain: EGF homology <EGF7>
F/1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 8.1%; Score 107; DB 2; Length 3707;
Best Local Similarity 21.7%; Pred. No. 24;
Matches 55; Conservative 39; Mismatches 95; Indels 64; Gaps 10;

```

```

DB 2661 PEKAVRNQRLRL 2673
RESULT 51
A38096
perlecan precursor - human
N/Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prot
C/Species: Homo sapiens (man)
C/Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 05-Nov-1999
C/Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R/Murdoch, A.D.; Bode, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A/Title: Primary structure of the human heparan sulfate proteoglycan from basement memb
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A/Reference number: A38096; MUID:92235084; PMID:1569102
A/Accession: A38096
A/Molecule type: mRNA
A/Residues: 1-4391 <MUR>
A/Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R/Kallunki, P.; Tyygvaasen, K.
J. Cell Biol. 116, 559-571, 1992
A/Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pr
all adhesion molecules, and epidermal growth factor.
A/Reference number: A41736; MUID:92112994; PMID:1730768
A/Accession: S19256
A/Molecule type: mRNA
A/Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908, 'R'
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-
A/Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA43373.1; PID:g29470
R/Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kesilae, M.; Shows, T.B.; Tyygvaasen, K.
Genomics 11, 389-396, 1991
A/Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the
A/Reference number: A41059; MUID:92120660; PMID:1685141
A/Accession: A41059
A/Molecule type: mRNA
A/Residues: 'RT', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397 <KA2>
A/Cross-references: GB:S76436; NID:g243370; PIDN:AA21121.1; PID:g243371
R/Dodge, G.R.; Kovalesky, I.; Chu, M.L.; Haesell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A/Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellu
A/Reference number: A40306; MUID:91365376; PMID:1679749
A/Accession: A40306
A/Molecule type: mRNA
A/Residues: 1018-1405, 'G', 1407-1409, 'G', 1411-1465 <DOD>
A/Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R/Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Caestejan, J.J.; van de
J. Cell Biol. 109, 3199-3211, 1989
A/Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A/Reference number: A33625; MUID:90078352; PMID:2687294
A/Accession: B33625
A/Molecule type: protein
A/Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
A/Accession: A33625
A/Molecule type: protein
A/Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185 <HE3>
A/Note: peptide potentially matches four different regions of sequence shown
C/Genetics:
A/Gene: GDB:HSRG2
A/Cross-references: GDB:126372; OMIM:142461
A/Map position: 1936.1-1936.1
C/Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G rep
C/Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembr
F/1-21/Domain: signal sequence #status predicted <SIG>

```

F:22-4391/Product: perlecan #status predicted <MAT>  
 F:22-193/Domain: I <DOM1>  
 F:194-530/Domain: II <DOM2>  
 F:199-534/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:531-1676/Domain: III <DOM3>  
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>  
 F:163-1610/Domain: laminin-type EGF-like homology <EGF7>  
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>  
 F:1677-3686/Domain: IV <DOM4>  
 F:2007-2034/Domain: transmembrane #status predicted <TRM>  
 F:3687-4391/Domain: V <DOM5>  
 F:3845-3880/Domain: EGF homology <EGF1>  
 F:3888-3921/Domain: EGF homology <EGF>  
 F:3953-4106/Domain: laminin G repeat homology <LG2>  
 F:4147-4175/Domain: EGF homology <EGF2>  
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif  
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif  
 F:65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted  
 F:89,554,1155,1212,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 8.1%; Score 107; DB 2; Length 4391;

Best Local Similarity 19.8%; Pred. No. 30; Indels 146; Gaps 11;

Matches 67; Conservative 31; Mismatches 94; Indels 146; Gaps 11;

QY 12 LVQLALLPATQGNKVVLTCTASQKSIQPHWKNNOIKILGNQSFLLTK 71  
 DB 3291 IILHVESPYATVPEHNASVQAGETVQLCLAGTPTPLTQW-----SR 3334  
 QY 72 GPKSLNDRADSRSLMDQGNPPLIKLKIEDSDTYICEVDD---KEEVQLVFG-- 124  
 DB 3335 VGSSLPRATARNEL-----LHFEAPABDSGRVCRVTNKGSAEAPQLLVQGP 3387  
 QY 125 -LTRANS-----DTHLLQGSLTLTLE-----SPGSSPS 152  
 DB 3388 SLPATSLPAGSTPTVQVTPQLTKTSIGASVDFHCAPVSDGQTQLRWKEGGQLPFGHSVQ 3447  
 QY 153 -----VQCRSPRGKN----- 162  
 DB 3448 DGVLRIGNLDQSCGYTICQAHGPMGKAKQASQALVIALPSVLINIRTSVQTVVGHAVE 3507  
 QY 163 -----IQGKTLVSQLELQDSGTWTCTYLQ-----NQK 192  
 DB 3508 FECLALDPRQVYTSKVGHLRPGIVQSGGVARIHAHELADAGQVRCCTNNAAGTQSH 3567  
 QY 193 VEFKIDIVPRASALPAP---PTGSALPDPPOTASALPP 227  
 DB 3568 VLLIVQALPQIS-MPQEVPRVAGSAAVFPCLASGTPFP 3604

# RESULT 52

B37266

Ig kappa chain V region (2G8) - mouse

C/Species: Mus musculus (house mouse)

C/Date: 18-Oct-1991 #sequence\_revision 18-Oct-1991 #text\_change 21-Jan-2000

C/Accession: B37266

R/Ruff-Jamison, S.; Campos-Gonzalez, R.; Glenney Jr., J.R.

J. Biol. Chem. 266, 6607-6613, 1991

A/Title: Heavy and light chain variable region sequences and antibody properties of anti

A/Reference number: A38740; MUID:91177923; PMID:11706720

A/Accession: B37266

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-111 <RUF>

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F:19-93/Domain: immunoglobulin homology <IMM>

Query Match 8.1%; Score 106.5; DB 2; Length 111;

Best Local Similarity 35.2%; Pred. No. 0.42;

Matches 32; Conservative 10; Mismatches 40; Indels 9; Gaps 3;

QY 23 TQGNKVVLTGKGDVTELTCTASQKSIQPHW---KNSNOIKILGNQSFLLTKPSKLNDR 79  
 DB 8 TQSTKMTSIVSDRVSITCKASQDVSNNYLMWQOKPDGTVKILYSASYSRTG---VPDR 64  
 QY 80 ADSRSLMDQGNPPLIKLKIEDSDTYICE 110  
 DB 65 FTGRGSGTD---FSFTISVKAEDLAVYCE 92

# RESULT 53

I38740

Ig kappa chain V region (Py42) - mouse

C/Species: Mus musculus (house mouse)

C/Date: 18-Oct-1991 #sequence\_revision 18-Oct-1991 #text\_change 21-Jan-2000

C/Accession: I38740

R/Ruff-Jamison, S.; Campos-Gonzalez, R.; Glenney Jr., J.R.

J. Biol. Chem. 266, 6607-6613, 1991

A/Title: Heavy and light chain variable region sequences and antibody properties of anti

A/Reference number: A38740; MUID:91177923; PMID:11706720

A/Accession: I38740

A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr.

A/Molecule type: mRNA

A/Residues: 1-111 <RUF>

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F:19-93/Domain: immunoglobulin homology <IMM>

Query Match 8.1%; Score 106.5; DB 2; Length 111;

Best Local Similarity 35.2%; Pred. No. 0.42;

Matches 32; Conservative 11; Mismatches 39; Indels 9; Gaps 3;

QY 23 TQGNKVVLTGKGDVTELTCTASQKSIQPHW---KNSNOIKILGNQSFLLTKPSKLNDR 79  
 DB 8 TQSTKMTSIVSDRVSITCKASQDVSNNYLMWQOKPDGTVKILYSASYSRTG---VPDR 64  
 QY 80 ADSRSLMDQGNPPLIKLKIEDSDTYICE 110  
 DB 65 FTGRGSGTD---FSFTISVKAEDLAVYCE 92

# RESULT 54

JC1239

Opioid-binding protein (clones S68 and SG13) - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 19-May-2000

C/Accession: JC1239

R/Lippman, D.A.; Lee, N.M.; Loh, H.H.

Gene 117, 249-254, 1992

A/Title: Opioid-binding cell adhesion molecule (OBAM)-related clones from a rat brain

A/Reference number: JC1238; MUID:92347701; PMID:1339369

A/Accession: JC1239

A/Molecule type: mRNA

A/Residues: 1-345 <LIP>

C/Cross-references: GB:M88710; NID:G203247; PIDN:AAA0859.1; PID:G203248; GB:M88711; NID

C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi

C/Keywords: transmembrane protein

Query Match 8.1%; Score 106.5; DB 2; Length 345;

Best Local Similarity 22.1%; Pred. No. 1.6;

Matches 58; Conservative 43; Mismatches 101; Indels 61; Gaps 10;

QY 11 LVQLALLPATQGNKVVLTCTASQKSIQPHWKNNOIKILGNQSFLLTKPSKLNDR 57  
 DB 14 LVVSLRLFLVPTGVPRSGDAPTPKAMDNTVVRQGSATLRCCTIDRVT-RVAMLNRS 72  
 QY 58 QIKILN-----QGSFLTKGPSKLNDRADSRSLMDQGNPPLIKLKIEDSDTYICE 110  
 DB 73 TIIYAGNDKMSIDPRVILVNTPTQ-----YIMIQNDVYDEGPTCS 116  
 QY 111 VE-----DKEEVQLVFG-----LTNSDTHLLQGSLTLTLESPPGSSPSVQCRSPRGKN 162

Db 117 VQDNHFKTSSVHLIVQVPQIMNISSDITVNEISSVTLCLALGPREPTVWHLVSK 176  
Qy 163 IQG-----GKTLVSQLEQDSGTWCTVQ-----NOKKVEFKIDIVPRASALPAPPG 213  
Db 177 GGGFVSEDEYLEISDIKDSGSEYCSALNDVAPDVKKITVNPYIS--KAKNTGV 234  
Qy 214 ALPDPQTASALPDPASALPAA 236  
Db 235 SVGOKGILSC-----EASAVPMA 252

## RESULT 55

134416  
hypothetical protein F12F3.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C/Accession: T34416  
R/Fulton, B.; Wohldmann, P.  
A/Description: The sequence of C. elegans cosmid F12F3.  
A/Reference number: 221521  
A/Accession: T34416  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-2783 <FULL>  
A/Cross-references: EMBL:U80022; PIDN:AAC25886.1; GSPDB:GN00023; CESP:F12F3.2  
C/Genetic8:  
A/Map position: 5  
A/Intons: 45/3; 90/3; 451/3; 509/1; 2313/3; 2341/3; 2378/3; 2414/2; 2453/3; 2474/2; 252

Query Match 8.1%; Score 106.5; DB 2; Length 2783;  
Best Local Similarity 22.6%; Pred. No. 19;  
Matches 56; Conservative 31; Mismatches 84; Indels 77; Gaps 10;  
Qy 28 VLLKSKDPTVELTCTASQKSIQFHWKNSNQIKILNGSFLTKGPKSLNDRADSRSLW 87  
Db 1187 VLLKTAGRTATFTQSYANPAQVW-----LHNGKALQOTKSNYKRLF 1231  
Qy 88 DQGNPFLIKNLKIEDSDTYICEVEDQEEV---QLLVFGLTANS----- 129  
Db 1232 DDMATVLIENVTDLCCTYAVANNQGDVHTSQTLISGEAKKIAASLPYFIELKP 1291  
Qy 130 DTHLQGSLLTLESPPGSSP-----SVQGRPR-----GKNIGGKTLSSQ 173  
Db 1292 KINVEG--ATLSIQADLNGSRIPEVWVKONSELVESDRIQMKCDGVNYG---LTVRD 1345  
Qy 174 LELQDSGTWCTCT-----VLQN-----QKVEFKIDIVPRASALPAPPGSA 214  
Db 1346 VGLEDEGTYITTAENEKIKINQNTVEVSTKSKVEKKEKKVEKDGKKKPGRP--G 1402  
Qy 215 LPDPQTAS 222  
Db 1403 LPRPGAS 1410

## RESULT 56

T43027  
neural cell adhesion molecule L1 - goldfish  
N/Alternate names: E587 antigen  
C/Species: Carassius auratus (goldfish)  
C/Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000  
C/Accession: T43027  
R/Giordano, S.; Leesing, U.; Lottspeich, F.; Stuenkel, C.A.O.  
A/Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule expressed in the brain.  
A/Reference number: 222294  
A/Accession: T43027  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1232 <GIO>  
A/Cross-references: EMBL:U55211; NID:G1305526; PID:G1305527; PIDN:AAA99159.1

C/Superfamily: neural cell adhesion molecule L1, fibronectin type III repeat homology;  
C/Keywords: cell adhesion; membrane protein

Query Match 8.0%; Score 106; DB 2; Length 1232;  
Best Local Similarity 23.2%; Pred. No. 7.8;  
Matches 54; Conservative 40; Mismatches 87; Indels 52; Gaps 11;  
Qy 2 NRGVPRFHLVLTQALLPAPQGNKVLGKGGDPTVELTCTASQKSIQFHWKNSNQIKI 61  
Db 286 SQGSVKHIVAVTVAA--PWTRPRPNHLVAGETVLDQAEIPFNITW--SMNGAPI 342  
Qy 62 LGNGSFLTKGPKSLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQEEVQLT 121  
Db 343 AGTD-----PPRRHV---SSGTLTLTVQI--SDTAVYHVEATNKHGNTL 383  
Qy 122 VFGLTNSDTHL--LQGSLLT--LTLESPPGSSPSVQCR--SPRGK----- 161  
Db 384 I-----NTHVNVVELPQIITLEDLKYEAEGQTVLLQCRFGSPQPRVWQITNSGPAL 438  
Qy 162 -NIGKGT---LSVQLQDSGTWCTVQGNQKVEFKIDIVPRASALPAP 209  
Db 439 ANAKMSQTSQNTLQISDVSEBSSMYTCSVSTMSISAEVLVNRKTYDPP 491

## RESULT 57

T08851  
Down syndrome cell adhesion protein 1 - human (fragment)  
N/Alternate names: Down syndrome cell adhesion molecule  
C/Species: Homo sapiens (man)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
R/Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenber  
submitted to the EMBL Data Library, September 1997  
A/Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down s  
A/Reference number: 216495  
A/Accession: T08851  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1896 <YAM>  
A/Cross-references: EMBL:AF023449; NID:G3169765; PID:G3169766  
A/Experimental source: brain; developmental stage: 14 weeks; fetal  
C/Genetics:  
A/Genes: DSCAM  
A/Map position: 21q22  
A/Note: derived from alternatively-spliced mRNA  
C/Function:  
A/Description: involved in nervous system development  
C/Keywords: alternative splicing

Query Match 8.0%; Score 106; DB 2; Length 1896;  
Best Local Similarity 21.5%; Pred. No. 13;  
Matches 50; Conservative 34; Mismatches 75; Indels 74; Gaps 10;  
Qy 34 GDTVELTCTASQKSIQFHWKNSNQIKILNGSFLTKGPKSLNDRADSRSLMDQGNF 93  
Db 225 GQVLELPCKALGHPEPRYRLKXNM-----PLELSGR-----FQKVTG 263  
Qy 94 LIINKLKIEDSDTYICEVEDQEEVQLLVFGLTNSDTHLLQGSLLTTL-----ESPFG 148  
Db 264 LTIENIRPSDGSVCEVSNR-----YG-TAKVIGRLVYQPLKATISPRKVSVG 314  
Qy 149 SSPVQCR-----SPRKNIQ-----GKTLVSQLEQDSGTWCTV 186  
Db 315 SQVLSGCVTGTEQDELQSWRNGEILNP--QKNVITGGINNENLIMDMVNSDGGAYOCFV 373  
Qy 187 LQNKVEFKIDIVPRASALPAPPGSALPD--PQTASALPDPASALPAA 237  
Db 374 RQDLASQDYQVYV-----LEDGTPKITISASEKVSFAPEVSL 412

## RESULT 58

149294  
CD7 antigen - mouse

C;Species: Mus musculus (house mouse)  
 C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jul-2000  
 C;Accession: I49294; I49588  
 R;Lee, D.M.; Watson, M.L.; Seldin, M.F.  
 Immunogenetics 39, 289-290, 1994  
 A;Title: Mouse Cd7 maps to chromosome 11.  
 A;Reference number: I49294; MUID:94164701; PMID:7509775  
 A;Accession: I49294  
 A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/EWE  
 A;Molecule type: DNA  
 A;Residues: 1-210 <RES>  
 A;Cross-references: EMBL:U23462; NID:9833815; PIDN:AB17482.1; PID:9833816  
 R;Yoshikawa, K.; Seto, M.; Ueda, R.; Obata, Y.; Fukatsu, H.; Segawa, A.; Takahashi, T.  
 Immunogenetics 37, 114-119, 1993  
 A;Title: Isolation and characterization of mouse CD7 cDNA.  
 A;Reference number: I49588; MUID:93138718; PMID:7678579  
 A;Accession: I49588  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-68, 'F', 70-210 <RES>  
 A;Cross-references: GB:D10329; NID:93868752; PIDN:BA01171.1; PID:9286071  
 C;Genetics:  
 A;Gene: Cd7  
 A;Intons: 25/1; 130/1; 175/3  
 C;Keywords: glycoprotein

Query Match 8.0%; Score 105.5; DB 2; Length 210;  
 Best Local Similarity 21.7%; Pred. No. 1;  
 Matches 57; Conservative 33; Mismatches 60; Indels 113; Gaps 13;  
 10 LLLVLQIA-LLPA-----ATGCGKVVLLGKKGDPVELTCTASQKKSIGPHMKNSNQIKL 62  
 Db 7 LALLLTLAGILPEGLADQVHQSRLLTASBGSVNITCSR----- 48  
 QY 63 GNQGSFLLTKGPKSLNDRADRSRLMDQGNFPLIKLIKIEDSPDYICEVEDQKEVQLV 122  
 Db 49 GHLEGLIMK-----KIMPCA-----YNYVLYEDRQE----- 74  
 QY 123 FGLTANSDTHLQGSLLTLESPPGSSPSVQCRSPRGK-NIQGK---TLSVSOLELD 178  
 Db 75 -----PTVD-RTFSGRINFSGSQKLLTITISLQAD 105  
 QY 179 SGTWCTVLONOKVE---FKIDIVPRASALPAPPGSALPDQTSALPDPPAAS-AL 233  
 Db 106 TGDYTCAV---KRVASRGLEFTTVVAVKES-----QEAIRSGEPLQTSFSF 149  
 QY 234 PALAVISFLIGLGLGVACVLAR 256  
 Db 150 PAALAVGFFFTGLLGVCSMLR 172

RESULT 59  
 PNO568  
 connectin 3B - chicken (fragment)  
 N;Alternate names: Cn3B protein  
 C;Species: Gallus gallus (chicken)  
 C;Date: 03-May-1994 #sequence\_revision 07-Oct-1994 #text\_change 17-Mar-1999  
 C;Accession: PNO568  
 R;Maruyama, K.; Endo, T.; Kume, H.; Kawamura, Y.; Kanazawa, N.; Nakauchi, Y.; Kimura, S.;  
 Biochem. Biophys. Res. Commun. 194, 1288-1291, 1993  
 A;Title: A novel domain sequence of connectin localized at the 1 band of skeletal muscle  
 A;Reference number: PNO568; MUID:93356802; PMID:8352787  
 A;Accession: PNO568  
 A;Molecule type: mRNA  
 A;Residues: 1-1323 <MAR>  
 A;Cross-references: DDBJ:D16541; NID:9391629; PID:d1004495; PID:9391630  
 A;Experimental source: skeletal muscle  
 C;Comment: This protein string-like single molecule spans from the 2 line to the M line

Query Match 8.0%; Score 105.5; DB 2; Length 123;  
 Best Local Similarity 19.5%; Pred. No. 9.3;  
 Matches 57; Conservative 45; Mismatches 98; Indels 93; Gaps 12;

QY 8 RHLLVLQIALLPAAATQGN-KVVLG-----KKGDTVELTCT 42  
 Db 928 RHLLKIKNCQL---EDQGYRIVCGPHIASALTYIEPVERHLDPTTKRGTTCTLSQ 964  
 QY 43 ASQKKSIOFHWKNSNOIKILGNQGSFLTGPSEKLNDRADRSRLMDQGNFPLIKLIKIE 102  
 Db 985 FSLPNAKSGWYNNGRIK---GGRYSTQVSKVH-----KLIKQVRE 1026  
 QY 103 DSDTYICEVED-----QKEBQVLVFGLTANSDTHLQGSLLTLESPPG----- 148  
 Db 1027 DQGYTCCKLDNLETADLTIEAEPIQF-----TYSIQNIWVSEHQSAFEECEVSPDDAV 1080  
 QY 149 ----SSPSVQCRSPRGKNIQGGKT---LSVSOLELDQSGTWCTC-----Y 186  
 Db 1081 VTWYKCPTELTLSPEKSPFSEGCWYTHNVTAEBGYSVIARLEPGEARSTAEVY 1140  
 QY 187 LQNKQKVEFKIDIVPRAS-ALPAPPGSALPDQTSALPD---PAASALP 234  
 Db 1141 VTKEIKLELKPDPVPDAKVAVPPQKPAEAPIPILLPLPTPEKKPAEKVP 1193

RESULT 60  
 T32735  
 telomerase-associated protein component 1 - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C;Accession: T32735  
 R;Nakayama, J.; Salto, M.; Nakamura, H.; Matsura, A.; Ishikawa, F.  
 Cell 88, 875-884, 1997  
 A;Title: TRP1: a gene encoding a protein component of mammalian telomerase is a novel m  
 A;Reference number: Z21217; MUID:97236507; PMID:9118230  
 A;Accession: T32735  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-2629 <NAR>  
 A;Cross-references: EMBL:U89282; NID:91932816; PID:g1932817; PIDN:AA851690.1  
 A;Experimental source: strain Fischer 344  
 C;Genetics:  
 A;Gene: TRP1  
 C;Function:  
 A;Description: modification of telomerase-associated protein component 1 may regulate t

Query Match 8.0%; Score 105; DB 2; Length 2629;  
 Best Local Similarity 21.7%; Pred. No. 23;  
 Matches 62; Conservative 40; Mismatches 82; Indels 102; Gaps 14;  
 9 HLLVLQIALLPAAATQGN-KVVLG-----KKGDTVELTCT 49  
 Db 2235 HPLLVQIHTLQHSQGPVTAASAASGILLTSDNSVRLWQIPKEADPTCKPRSSAVIT 2294  
 QY 50 QPHKNSNOIKILGNQGSFLT-----KGPSEKLNDRADRSRL 86  
 Db 2295 AVAMAPDGSILVSGNAGELTLMQKAQAVATAPARVSDLWCSANAFVLSANVSE 2354  
 QY 87 WD-----QG---NPLIKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGS 138  
 Db 2355 WQVELARKSGCTCNFRILYLRVLEDDG-----VLTMALAPD-----GQS 2394  
 QY 139 LTL-----TLESPPGSSPSVQCRSPRGKNIQGGKTLSVSOLELDQSD 179  
 Db 2395 LILMKEDVELQWKPSTPSSICRYAVAHSSILCTSKDGLFYLQCGNSGSLILEQES 2454  
 QY 180 GTWCTVLONOKKVEFKIDI-VPRASALPAPPGSALPDQTSAL 224  
 Db 2455 GKF-----EKTLDENLNLNPNNGS-----PVSTIQAEPSGSSL 2488

RESULT 61  
 A32579  
 neuroglian - fruit fly (Drosophila melanogaster)  
 C;Species: Drosophila melanogaster  
 C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C;Accession: A32579

R:Blieber, A.J.; Snow, P.M.; Hortsch, M.; Patel, N.H.; Jacobs, J.R.; Traquina, Z.R.; Schi Cell 59, 447-460, 1989  
 A:Title: Drosophila neuroglian: a member of the immunoglobulin superfamily with extensive  
 A:Reference number: A2579; MUID:90030418; PMID:2805067  
 A:Accession: A2579  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-123 <BIE>  
 A:Cross-references: GB:M8221; NID:G157998; PID:AAA28728.1; PID:G157999  
 A>Note: the authors translated the codon TAT for residue 1234 as Thr and AAA for residue  
 C:Genetics:  
 A:Gene: FlyBase:FlyBase:FBgn0002966  
 A:Cross-references: FlyBase:FBgn0002966  
 C:Superfamily: neural cell adhesion molecule 11; fibronectin type III repeat homology;  
 C:Keywords: alternative splicing; cell adhesion; duplication; membrane protein  
 F:353-412/Domain: immunoglobulin homology <IMM1>  
 F:446-502/Domain: immunoglobulin homology <IMM2>  
 F:535-596/Domain: immunoglobulin homology <IMM3>

Query Match 7.9%; Score 104.5; DB 1; Length 1239;  
 Best Local Similarity 20.4%; Pred. No. 10;  
 Matches 52; Conservative 36; Mismatches 106; Indels 61; Gaps 11;

Qy 25 GNKVVLGGKQDTVELTCTASQKSIQFMKNSNQIKIINGOGSFLTKGPSKLNDRADRR 84  
 Db 233 GNKVLLDSQNHPPRQVSRQSLALGKMEFLCIYGG-----TLPQVWMSK-DGCR 286  
 Qy 85 SLWD---QGNP--PLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLQGS 138  
 Db 287 IQMSDRITQGHYKSLVIRQTNFDPAGTYCTDVNSGVNAQSFIILNVNSVPEFTKEPE 346  
 Qy 139 LTLTLESPGSSPSVQGS-----PRGNKIOGK-----TLVSQLELQ 177  
 Db 347 IATAED--EEVVEECAPAGVPEPKISWINGKPIEOSTNPRRTVTDNTIRIINLVKG 403  
 Qy 178 DSGTWTCTVLQNOCKVEFKIDIVPRASALPAPPTGSALPDQTASALPDPPASALPAAL 237  
 Db 404 DTGNYGCA--INSLSYVVK-DVLYNVQA-----EPTTISEAPAAV 441  
 Qy 238 AVISFLGLGLGVAC 252  
 Db 442 STDV---GRNVTKC 453

## RESULT 62

JC1238  
 A:Title: Opioid-binding protein (clone DU21) - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 19-May-2000  
 C:Accession: JC1238  
 R:Lipman, D.A.; Lee, N.M.; Loh, H.H.  
 A:Title: Opioid-binding cell adhesion molecule (OBAM)-related clones from a rat brain c  
 A:Reference number: JC1238; MUID:92347701; PMID:1339369  
 A:Accession: JC1238  
 A:Molecule type: mRNA  
 A:Residues: 1-338 <LIP>  
 A:Cross-references: GB:M68709; NID:G203245; PID:AAA40858.1; PID:G203246  
 A:Experimental source: brain  
 C:Genetics:  
 A:Gene: OBAM  
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi

Query Match 7.9%; Score 104; DB 2; Length 338;  
 Best Local Similarity 22.6%; Pred. No. 2.4;  
 Matches 58; Conservative 41; Mismatches 102; Indels 56; Gaps 11;

Qy 4 GVPFRHLLVQLALPPAATGKNVYLGKGDYVELTCTASQKSIQFMKNSNQIKI 63  
 Db 21 GVPFRS-----GDAFFPAMDN--VTYRQGESATLRCTIDRVT-RVAMINRSTIYAG 71  
 Qy 64 N-----QGSFLTKGPSKLNDRADRRSLMDQGNFPLIKNLKIEDSDTYICEVE--- 112

Db 72 NDKMSIDPRVLIIVNTPTQ-----VSIMIQNVVDYDEGPYCSVQTDNH 115  
 Qy 113 DQKEVQLVYGV-----LTANSDFHLLQGSITLTLBSPGSSPSVQCRPRGKNIQ--- 165  
 Db 116 PKTSRVHLIVQVPPQINNISSDITVNEISSVTLLCLIGRPEPTVWTHSVKQGQFVS 175  
 Qy 166 -GKTLVSQLELDQSGTWTCTVLQ-----NOKVEFKIDIVPRASALPAPPTGSALPDPPQ 219  
 Db 176 EDLTLEISDIKRDGSGEYCSALNDVAAPDVRKKTIVNVPYIS--KAKTGVSVQKG 233  
 Qy 220 TASALPDPPASALPAA 236  
 Db 234 TLSC-----ASAVPMA 245

## RESULT 63

156551  
 A:Title: Rattus norvegicus (Norway rat)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 19-May-2000  
 C:Accession: 156551  
 R:Struyk, A.F.; Canoll, P.D.; Wolfgang, M.J.; Rosen, C.L.; D'Eustachio, P.; Salzer, J.I  
 J. Neurosci. 15, 2141-2156, 1995  
 A:Title: Cloning of neurotrophin defines a new subfamily of differentially expressed ne  
 A:Reference number: 156551; MUID:95198094; PMID:7891157  
 A:Accession: 156551  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-344 <RES>  
 A:Cross-references: EMBL:U16845; NID:G755184; PID:AAA67445.1; PID:G755185  
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi

Query Match 7.9%; Score 104; DB 2; Length 344;  
 Best Local Similarity 20.8%; Pred. No. 2.4;  
 Matches 53; Conservative 47; Mismatches 109; Indels 46; Gaps 9;

Qy 11 LTVQLALPPAATGKNVYLG-----KKGDYVELTCTASQKSIQFMKNSN 57  
 Db 14 LVVSLRLFLVPPGVVRSQDATTFRKAMDNVYRQGESATLRCTIDNRVT-RVAMINRS 72  
 Qy 58 QIKILNQGSFLTKGPSKLNDRADRRSLMDQGNFPLIKNLKIEDSDTYICEVEDQEE 117  
 Db 73 TLVAGNDKMKCLDPRVLLSN-----TQGYSEIQVDVYDGPYCSVQTDNHP 123  
 Qy 118 -----VOLVFGLTANSDFHLLQGSITLTLBSPGSSPSVQCR--SPRKN-IQGG 166  
 Db 124 KTSRVHLIVQVSPKIVISSDISINEGNISLTICATGRPEPTVWTHISPKAVFVSED 183  
 Qy 167 KTLVSQLELDQSGTWTCTVLQ-----QKVEFKIDIVPRASALPAPPTGSALPDQTA 221  
 Db 184 EYLETIGITRQSESEYCSASNDVAAPVVRVNVTVNPPYIS---EAKTGVVQKG 239  
 Qy 222 SALPDPPASALPAA 236  
 Db 240 TLQGE---ASAVPMA 251

## RESULT 64

JC1509  
 A:Title: biliary glycoprotein E - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 07-May-1999  
 C:Accession: JC1509  
 R:McCuag, K.; Rosenberg, M.; Nedellec, P.; Turbide, C.; Beauchemin, N.  
 Gene 127, 173-183, 1993  
 A:Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopr  
 A:Reference number: JC1505; MUID:93273228; PMID:8500759  
 A:Accession: JC1509  
 A:Molecule type: mRNA  
 A:Residues: 1-458 <MCC>  
 A:Cross-references: GB:X67280  
 C:Comment: This protein is expressed at the cell surface and plays a determinant role i

[illegible]

Db 409 LALAPTEENPMKKKILAAAGCRVILIECKKAPKPFWSW-----SKGT 452

Qy 74 SKLNDRADSRSLWDGQNFPLIILKNLIKEDSDPIICEVEQKEVQLLVFGLTANSDTHL 133

Db 453 EWLVN--SSRLIWEQDS--LEINNIIRNDGIGYTTCFAENNRK-----ANS---- 495

Qy 134 LOGQSLLTLESPP-----GSSPSVQCRS----- 157

Db 496 ----TGLVITDPFRILLAPINADITVGENATMQCAASFPDADLTFTWSENGYVIDFNK 551

Qy 158 ----PRGKNIQGGKTLVSQLELDQSGTWTCYVLQNGKVEFKIDIVPRASALPAPPTG 212

Db 552 ENIHQGNFPLDSNGELLIRNAQLKHAGRYCTAQTIVDNSSASADLVVRGP--PGPG 609

Qy 213 SALPDPQTAS 222

Db 610 IRIEDIRATS 619

RESULT 66

S05944

neuronal cell surface protein F3 precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 21-Jan-2000

C:Accession: S05944

R:Gennarini, G.; Cibelli, G.; Rougon, G.; Mattei, M.G.; Goridis, C.

J: Cell Biol. 109, 775-788, 1989

A:Title: The mouse neuronal cell surface protein F3: a phosphatidylinositol-anchored mem

A:Reference number: S05944; MUID:89340657; PMID:2474555

A:Accession: S05944

A:Molecule type: mRNA

A:Residues: 1-1020 <SEN>

A:Cross-references: EMBL:X14943; NID:950937; PIDN:CAA33075.1; PID:950938

C:Genetics:

A:Map position: 15F

C:Superfamily: contacting; fibronectin type III repeat homology; immunoglobulin homology

F:1-20/Domain: signal sequence #status predicted <SIG>

F:256-312/Domain: immunoglobulin homology <IMM>

Query Match 7.9%; Score 104; DB 2; Length 1020;

Best local similarity 20.2%; Pred. No. 8.8;

Matches 51; Conservative 28; Mismatches 89; Indels 84; Gaps 9;

Qy 16 LALPAPATQG--NIVVLGKKGDVTELTCTASQKKSIOFHMKNNSQIKLNGSFLTKGP 73

Db 409 LALAPTEENPMKKKILAAAGCRVILIECKKAPKPFWSW-----SKGT 452

Qy 74 SKLNDRADSRSLWDGQNFPLIILKNLIKEDSDPIICEVEQKEVQLLVFGLTANSDTHL 133

Db 453 EWLVN--SSRLIWEQDS--LEINNIIRNDGIGYTTCFAENNRK-----ANS---- 495

Qy 134 LOGQSLLTLESPP-----GSSPSVQCRS----- 157

Db 496 ----TGLVITDPFRILLAPINADITVGENATMQCAASFPDADLTFTWSENGYVIDFNK 551

Qy 158 ----PRGKNIQGGKTLVSQLELDQSGTWTCYVLQNGKVEFKIDIVPRASALPAPPTG 210

Db 552 ENIHQGNFPLDSNGELLIRNAQLKHAGRYCTAQTIVDNSSASADLVVRGP--PGPG 609

Qy 211 TGSALPDPQTAS 222

Db 610 GGLRIEDIRATS 621

RESULT 67

A57112

contactin precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jan-2000

C:Accession: A57112

R:Pejles, E.; Nativ, M.; Campbell, P.L.; Sakurai, T.; Martinez, R.; Lev, S.; Clary, D.O.

J: Cell 82, 251-260, 1995

A:Title: The carbonic anhydrase domain of receptor tyrosine phosphatase beta is a funct



A:Reference number: A57112; MUID:95354206; PMID:7628014  
 A:Accession: A57112  
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tree  
 A:Molecule type: mRNA  
 A:Residues: 1-1021 <PEL>  
 C:Superfamily: connectin; fibronectin type III repeat homology; immunoglobulin homology  
 C:Keywords: membrane protein; phosphatidylinositol linkage  
 F:256-312/Domain: immunoglobulin homology <IMM>

Query Match 7.9%; Score 104; DB 2; Length 1021;  
 Best Local Similarity 20.2%; Pred. No. 8.8;  
 Matches 51; Conservative 28; Mismatches 89; Indels 84; Gaps 9;

16 LALLPAATQG--NKVVLKGGDTVELTCTASQKSIQFHWNKSNQIKILNGSFLTKGCP 73  
 DB LALLPFTFEMNPMKKILAKGRVIECKPKAAPKPKFSW-----SKGT 452  
 74 SKLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEVQLVFGILTANSDTHL 133  
 DB EMLVW--SSRLIWMEDGS--LEINNTIRNDGITYCFENNRK-----ANS----- 495  
 134 LQGSGLTTLSSPP-----GSSPSVQGRS----- 157  
 DB 496 ----TGLVITNPFTIILAPINADITVGENATMCCASFPDPLDTFVWSFNGYVIDENK 551  
 158 -----PRGNIOGKTLVSQLELDQSGTMTCTVLONOKKVEKIDIVRASALPAP 210  
 DB 552 EITHIHYQNFMLDANGELLIRNAQLKHAGRYCTTAQTTVDNSSASADLVVRGP--FGPP 609  
 211 TGSALPDPQTAS 222  
 DB 610 GGLREDIRATS 621

## RESULT 68

187731  
 Ig kappa chain V-J region (MSI-N17) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 05-Dec-1998 #text\_change 23-Jul-1999  
 C:Accession: S18731  
 R:Hitama, T.; Takehita, S.; Yoshida, Y.; Yamagishi, H.  
 Immunol. Lett. 27, 19-24, 1991  
 A:Title: Structure of extrachromosomal circular DNA generated by immunoglobulin light c  
 A:Reference number: S18731; MUID:91209891; PMID:1902191  
 A:Accession: S18731  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-103 <HIR>  
 A:Cross-references: EMBL:X54753; NID:955294; PION:CAA38555.1; PID:955295  
 A:Experimental source: spleen, strain BALB/C-nu/nu  
 A>Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1990  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka  
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 18  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:11-85/Domain: immunoglobulin homology <IMM>  
 F:18-83/Disulfide bonds: #status predicted

Query Match 7.9%; Score 103.5; DB 2; Length 103;  
 Best Local Similarity 39.0%; Pred. No. 0.63;  
 Matches 30; Conservative 5; Mismatches 37; Indels 5; Gaps 2;

34 GDTVELTCTASQKSIQFHWNKSNQIKILNGSFLTKGSPKLNDRADSRSLMDQ--NF 92  
 DB LGLVLTMTQASQGSINLWFOQKP---GKAPKLLIYGASNLIEDGVPSRSGSRYGTF 66  
 93 PLIKNLKIEDSDTYIC 109  
 DB 67 TLTISSLEDEDMATYFC 83

## RESULT 69

KVMS06

Ig kappa chain V region (J606) - mouse (tentative sequence)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 31-Mar-2000  
 C:Accession: A92811; A01929  
 R:Johnson, N.; Stankard, J.; Paul, L.; Hood, L.  
 J. Immunol. 128, 302-307, 1982  
 A:Title: The complete V domain amino acid sequences of two myeloma inulin-binding prote  
 A:Reference number: A92811; MUID:8209361; PMID:6798111  
 A:Accession: A92811  
 A:Molecule type: protein  
 A:Residues: 1-108 <JCH>  
 C:Comment: This chain was isolated from a myeloma protein that binds inulin.  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka  
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:16-90/Domain: immunoglobulin homology <IMM>  
 F:23-88/Disulfide bonds: #status predicted

Query Match 7.9%; Score 103.5; DB 1; Length 108;  
 Best Local Similarity 39.0%; Pred. No. 0.67;  
 Matches 30; Conservative 5; Mismatches 37; Indels 5; Gaps 2;

34 GDTVELTCTASQKSIQFHWNKSNQIKILNGSFLTKGSPKLNDRADSRSLMDQ--NF 92  
 DB LGLVLTMTQASQGSINLWFOQKP---GKAPKLLIYGASNLIEDGVPSRSGSRYGTF 71  
 93 PLIKNLKIEDSDTYIC 109  
 DB 72 TLTISSLEDEDMATYFC 88

## RESULT 70

521668  
 Ig kappa chain V region (24) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Jun-2000  
 C:Accession: S21668  
 R:Rooschenhaler, F.; Schaeble, K.F.; Thiebe, R.; Zachau, H.G.  
 Biol. Chem. Hoppe-Seyler 373, 177-186, 1992  
 A:Title: Of orphons and UHOB. Delimitation of the germline repertoire of human immunogl  
 A:Reference number: S21666; MUID:92281681; PMID:1596359  
 A:Accession: S21668  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-117 <ROE>  
 A:Cross-references: EMBL:X64642; NID:9432651; PIDN:CAA45916.1; PID:91335151  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:38-112/Domain: immunoglobulin homology <IMM>

Query Match 7.9%; Score 103.5; DB 2; Length 117;  
 Best Local Similarity 33.0%; Pred. No. 0.74;  
 Matches 37; Conservative 9; Mismatches 51; Indels 15; Gaps 4;

11 LVLVQLLPPA-----TQNKVVLKGGDTVELTCTASQKSIQFHWNKSNQIKILGN 64  
 DB LGLVLTMTQASQGSINLWFOQKP---GKAPKLLIYGASNLIEDGVPSRSGSRYGTF 64  
 65 QGSLTGTGSPKLNDRADSRSLMDQ--NFPLIKLKIEDSDTYICEVED 113  
 DB 65 TPKEFLTYAASLSLQGSIRFS--DSGSGTDTLTISLQFEDPATYICQOQSD 114

## RESULT 71

KJHWK  
 Ig kappa chain precursor V-I region (Walker) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 21-Jan-2000  
 C:Accession: A01883  
 R:Klobeck, H.G.; Combratio, G.; Zachau, H.G.  
 Nucleic Acids Res. 12, 6995-7006, 1984  
 A:Title: Immunoglobulin genes of the kappa light chain type from two human lymphoid cel

A:Reference number: A93534; MUID:85014148; PMID:6091049  
 A:Accession: A01883  
 A:Molecule type: DNA  
 A:Residues: 1-129 <KLO>  
 A>Note: the sequence was determined from the differentiated gene  
 C:Gene: GDB:IGKV1  
 A:Cross-references: GDB:136264  
 A:Map position: 2p12-2p12  
 C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:1-22/Domain: signal sequence #status predicted <Sig>  
 F:23-129/Product: Ig kappa chain V-I region (Walker) #status predicted <MAT>  
 F:23-45/Region: framework 1  
 F:38-112/Domain: immunoglobulin homology <IMM>  
 F:46-56/Region: complementarity-determining 1  
 F:57-71/Region: framework 2  
 F:72-78/Region: complementarity-determining 2  
 F:79-110/Region: framework 3  
 F:111-119/Region: complementarity-determining 3  
 F:120-129/Region: framework 4  
 F:45-110/Disulfide bonds: #status predicted

Query Match 7.9%; Score 103.5; DB 1; Length 129;  
 Best Local Similarity 32.8%; Pred. No. 0.83;  
 Matches 43; Conservative 10; Mismatches 61; Indels 17; Gaps 5;

Db 1 MNRGVPFRLHLLVQLALPAA-----TGQNKVVLGKGDVTELTCTASOKSIOPHWK 54  
 1 MDRKVPKQ--LIGILLMLKAGRCIDIQTSPSSLSVGDRTITRAQGISNLIWY 58

QY 55 NSNOIKILNGSGFLTKGPKLNDRADSRSLMDQG--NPLIILKLIKEDSDTYICEVD 113

Db 59 QQKP-----GKAPKLLIYAASSLGCVTSRFGSGSDTFLITSLQPEDSATYVC---- 110

QY 114 QKEVQLLVFG 124

Db 111 QQSYSTLITFG 121

RESULT 72

A53449  
 plasmacytoma-associated neuronal glycoprotein PANG - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 25-Aug-1995 #sequence\_revision 25-Aug-1995 #text\_change 24-Sep-1999  
 C:Accession: A53449  
 R:Connolly, M.A.; Grady, R.C.; Mushinski, J.F.; Marcu, K.B.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 1337-1341, 1994  
 A:Title: PANG, a gene encoding a neuronal glycoprotein, is ectopically activated by intr  
 A:Reference number: A53449; MUID:94151325; PMID:8108413  
 A:Accession: A53449  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1028 <CON>  
 A:Cross-references: GB:L01991; NID:g200056; PIDN:AAAI7403.1; PID:g200057  
 C:Superfamily: connectin; fibronectin type III repeat homology; immunoglobulin homology  
 C:Keywords: glycoprotein

Query Match 7.9%; Score 103.5; DB 2; Length 1028;  
 Best Local Similarity 22.1%; Pred. No. 9.6;  
 Matches 53; Conservative 33; Mismatches 83; Indels 71; Gaps 10;

QY 23 TQGNKVVLGKK-----GDTVELTCTASOKK--SIQPHKNSNOIKILNGSGFLTK-GP 73  
 497 TEPRITILASNMVAVAGESVILPCQVQHPDLIDIMFAMVFNGLTDFKDGSHFEKVG 556

Db 74 SKLNDRADSRSLMDQGNPLIILKLIKEDSDTYICEVDQEKERV-----QLLVFG----- 124

QY 557 SSSGD-----LMIRNIOLKISGKVVCMVQGVTSVSAELIYVGSFGPP 601

QY 125 -----LTANSDTHLLOGQSLTLTLESPPSSPVSQCRSPRKNKIQGCKT 168

Db 602 ENVKDEITDTTQASWTBGTDSH-----SEVISYAVQARFPFSGVQSVRT 648

QY 169 LSVQLELODQSGTWCTCTVLQONKVEFKDIDVPRASALPAPPGSALPDP--QTASALPD 226  
 649 VP-----EVIDGKTHRTATVVELNPNWVEYEFRIYASNKIGGEP---SLPSEKRTBEADE 701

RESULT 73

T20992  
 hypothetical protein F15G9.4a - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T20992; T24733  
 R:Subleat, J.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: Z19355  
 A:Accession: T20992  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5175 <W1>  
 A:Cross-references: EMBL:247068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a  
 A:Experimental source: clone F15G9  
 R:Kershaw, J.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: Z19929  
 A:Accession: T24733  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5175 <W12>  
 A:Cross-references: EMBL:247070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a  
 C:Experimental source: clone T0989  
 C:Gene: CESP:F15G9.4a  
 A:Map position: X  
 A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4878/1; 4941/1; 5011/1; 5077/1;

Query Match 7.9%; Score 103.5; DB 2; Length 5175;  
 Best Local Similarity 18.7%; Pred. No. 65;  
 Matches 62; Conservative 50; Mismatches 108; Indels 111; Gaps 16;

QY 26 NKVVLGKGDVTELTCTASOKSIOPHW-----KSNQIKILG-----NQ 65  
 709 NKQWVG-RGDRVSEFCKTIRGPHPIRPFKNGKDIKPDYIKINEGOLHWGABEDA 767

QY 66 GSF-----LTGPKSLNDRADSRSL-----MD 88

Db 768 GAVSCGEMMAGKDVQVAVLSVGRVPTTIESHTVAVNIEROTVLOCLAVGIRPPEIENG 827

QY 89 QGNF-----PLIILKLIKEDSDTYICEVD-----QKEVQLLVFGTLTA- 127  
 828 KGNVLLATINPRYTQLADGNLITDAQIEDQGFCTIRANRYGQOSGSTLLMTVLVSP 887

QY 128 -----NSDHLLOGQSLTLT-----LESPGSSPSVQCRSP--RGKNIQ---GGKTLVSQ 173

Db 888 VLGHPPEBQLLEGQDLTSLSCVVLGTSPSIWIMDKRVEBPTIKIEGGSLRLKLG 947

QY 174 LELQDSGTWCTCTVLQ--NOKVVEFKIDIVPRASALPAPPGSALPDPQTASALPDP 231  
 948 GNRKDGKCTCTCLAVSANGSTLHINVLKKEBFYVKKPEGIVF-----KPTIS 996

QY 232 ALPAA-LAVIS-----FLGLGLGVACVLART 257

Db 997 GMDKHAVALVNSTHVDVLDEGFAIPCVVSGT 1027

RESULT 74

T43290  
 hemiscentin precursor - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000

C:Accession: T43290; T20993; T24734  
 R:Vogel, B.E.; Hedgecock, E.M.  
 Submitted to the EMBL Data Library, June 1998  
 A:Description: Hemiscentin is required for hemidesmosome mediated cell adhesion and germ-  
 A:Reference number: 222396  
 A:Accession: T43290  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-5198 <VOG>  
 A:Cross-references: EMBL:AF074901; PIDN:AA026792.1  
 R:Subston, J.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: 219355  
 A:Accession: T20993  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5198 <WLL>  
 A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:FL5G9.4b  
 R:Kershaw, J.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: 219929  
 A:Accession: T24734  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5198 <WT2>  
 A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:FL5G9.4b  
 A:Experimental source: clone T0989  
 C:Genetic8:  
 A:Gene: hlm-4; FL5G9.4b  
 A:Map position: 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
 A:Insertions: 85/1; 120/1; 259/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
 A:Deletions: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
 A:4225/1; 4361/1; 4408/1; 4456/1; 4458/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/

Query Match 7.8%; Score 103.5; DB 2; Length 5198;  
 Best Local Similarity 18.7%; Pred. No. 65;  
 Matches 62; Conservative 50; Mismatches 108; Indels 111; Gaps 16;

QY 26 NKVVLLGKKGTVELTLC-TASQKKSIOFHW-----KNSNQIKILG---NQ 65  
 DB 709 NKMVVG-RGDRVSFECKTIRGKPHKIRFWKNGKDLIKPDYIKINEGQILIMGAKDEDA 767  
 QY 66 GSF-----LTGPKSLNDRADSRSL-----WD 88  
 DB 768 GAYSCVGENMAGKDVQVANISVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEMO 827  
 QY 89 QGNF-----PLIINKLKIEDSDTYICEVED---OKEEVQLVFGTLA- 127  
 DB 828 KGNVLLATLNPNRYTQQLADGNLITLTDQIEDQOFTCIARTYGOQOSQSTLMTGLVSP 887  
 QY 128 ----NSDTHLLQGQSLTLT---LESPPGSSPSVQCRSP--RKNKIQ--GKTLVSQ 173  
 DB 888 VLGHVPEEQQLIEGQDLTLSCVVVLGTPKPSIWKDKDPVEBPPTIKIEGGGSLRLRG 947  
 QY 174 LELODSGWTCTVQ--NQKVERKIDIVPRASALPAPPTGSALPDQTASALDPDPAAS 231  
 DB 948 GNPDEGXYTCIAVSPAGNSTLHINVOLIKKPEFYVKEGGIVF-----KPTIS 996  
 QY 232 ALPAA-LAVIS-----FLTGLGLGAVCVLART 257  
 DB 997 GMDKHYAVVNSTHVDLDEGFAIPCWVSGT 1027

RESULT 75  
 A44783  
 ecto-ATPase precursor - rat  
 N:Alternate names: canalicular bile acid transport protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 23-Jul-1999  
 C:Accession: A44783; A44110  
 R:Lin, S.H.; Guidotti, G.  
 J. Biol. Chem. 264, 14408-14414, 1989

A:Title: Cloning and expression of a cDNA coding for a rat liver plasma membrane ecto-A  
 A:Reference number: A44783; MUID:89340561; PMID:2527235  
 A:Accession: A44783  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-519 <LIN>  
 A:Cross-references: GB:04963; NID:9203969; PIDN:AAA41104.1; PID:9203990  
 R:Sippel, C.J.; Suchy, F.J.; Nanthanarayanan, M.; Perlmuter, D.H.  
 J. Biol. Chem. 268, 2083-2091, 1993  
 A:Title: The rat liver ecto-ATPase is also a canalicular bile acid transport protein.  
 A:Reference number: A44110; MUID:93131966; PMID:8420979  
 A:Accession: A44110  
 A:Molecule type: Protein  
 A:Residues: 110-120;122-138;148-150 <STP>  
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi  
 C:Keywords: glycoprotein; liver; phosphoprotein; transmembrane protein  
 F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>  
 F:160-217/Domain: immunoglobulin homology <IMM1>  
 F:252-301/Domain: immunoglobulin homology <IMM2>  
 F:337-394/Domain: immunoglobulin homology <IMM3>

Query Match 7.8%; Score 103; DB 2; Length 519;  
 Best Local Similarity 20.9%; Pred. No. 47;  
 Matches 51; Conservative 34; Mismatches 93; Indels 66; Gaps 7;

QY 33 KGPVELTCTASQKKSIOFHWKNSNQIKILGNGSFLTKPSKLNDRADSRSLMDQGNF 92  
 DB 251 QGSNLTNSCHADSNPPAQYFWLNEKQTSQSE----- 283  
 QY 93 PLIINKLKIEDSDTYICEVEDOKEEVQLVFGTLAN-----SDTHLL 134  
 DB 284 -LFTSNTTNNSGTYACFVNN-----TVTGLSRTYKNIITVEPPTQPSIQITNTTVK 335  
 QY 135 QGQSLTTLTSPG-----SSPSVQCRSPKKNIQGKTLVSQLELODSGWTCTV 186  
 DB 336 ELGSVTLTCSKDTGVSVRWLFNSQSLQ-TDWMTLSDQNSTLRIDPIKREDAGDYQCEI 394  
 QY 187 LQNGKVEKIDIVPRASALPAPPTG-SALPDQTSALDPDPAASALPAAVLVISFLG 245  
 DB 395 ---SNPFSRISHPIKLDVIPDPQTGNSGSEGAIGIVGSVAGVALIAPLAVLSRK 451  
 QY 246 LGIG 249  
 DB 452 TGGG 455

RESULT 76  
 149583  
 differentiation antigen - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
 C:Accession: 149583  
 R:Law, C.  
 J. Immunol. 151, 175-187, 1993  
 A:Title: Organization of the murine Cd22 locus. Mapping to chromosome 7 and characteriz  
 A:Reference number: 149583; MUID:93315834; PMID:8100643  
 A:Accession: 149583  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-862 <RES>  
 A:Cross-references: GB:116928; NID:9348965; PIDN:AAA02562.1; PID:9348966  
 C:Genetic8:  
 A:Gene: CD22

Query Match 7.8%; Score 102.5; DB 2; Length 862;  
 Best Local Similarity 21.9%; Pred. No. 92;  
 Matches 52; Conservative 33; Mismatches 85; Indels 67; Gaps 10;

QY 32 KGPVELTCTASQKKSIOFHWKNSNQIKILGNGSFLTKPSKLNDRADSRSLMDQ 89  
 DB 535 RAGQRYVLQCDFAESNPAEVRFFWKNGSLVQGRYLSFGSVSPE-----DS 581  
 QY 90 GNPFLIINKLKIEDSDTYICEVEDOKEEVQLV-----FGLTANSDFHLLQGQSLTTLTSS 145

```

Db      582 GATNCWNN-----SIEETISQAMNLOVLYAPRRRLVSIPIGDHVEGKATLSCS 633
QY      146 ---PPGSSPSVQCRSPRGKNI-OQGKTLVSQLELODSGTWTCTVLONOKKVEFKDIYV 201
Db      634 DAPPIISQYWPDS--GQDLHSSGQKLRLEPLEVQHTGSYRC-----674

QY      202 RASALPAPPTGSAIPDQTASALPDPASALPAALAVISFLGLGIG--VACVIA 255
Db      675 --KGTNGIGTGESPPSTLTLYYSPE-----TIKRYALGIGFCLTITCIIA 717

RESULT 77
S51556
Vascular endothelial growth factor receptor 1 precursor - Japanese quail
N:Alternate names: quail endothelial kinase 1 (Quek 1); vascular endothelial growth fac
C:Species: Coturnix coturnix japonica (Japanese quail)
C>Date: 07-May-1995 #sequence revision 01-Sep-1995 #text_change 08-Dec-2000
C:Accession: J04953; A56598; T51162; S51656
R:Etchmann, A.; Marcelle, C.; Breant, C.; Le Douarin, N.M.
Gene 174, 3-8, 1996
A:Title: Molecular cloning of Quek 1 and 2, two quail vascular endothelial growth factor
A:Reference number: J04953; MUID:97017121; PMID:8863722
A:Accession: J04953
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1348 <E1C1>
A:Cross-references: EMBL:X83288; NID:G603523; PIDN:CAA58268.1; PID:G603524
A:Note: submitted to the EMBL Data Library, December 1994
R:Etchmann, A.; Marcelle, C.; Breant, C.; Le Douarin, N.M.
Mech. Dev. 42, 33-48, 1993
A:Title: Two molecules related to the VEGF receptor are expressed in early endothelial c
A:Reference number: A56598; MUID:93378866; PMID:8336413
A:Contents: E16 spinal cord
A:Accession: A56598
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 310-1348 <E1C2>
A:Cross-references: GB:S65205; NID:G410680; PIDN:AA828127.1; PID:G410681
A:Note: sequence extracted from NCBI backbone (NCBIN:137162, NCBIPI:137163)
R:Marcelle, C.; Etchmann, A.
Oncogene 7, 2479-2487, 1992
A:Title: Molecular cloning of a family of protein kinase genes expressed in the avian en
A:Reference number: I50595; MUID:93096482; PMID:1281306
A:Accession: I51162
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1023-1079 <MAR>
A:Cross-references: EMBL:X69694; NID:G395226; PIDN:CAA49364.1; PID:G3938278
A:Note: the species is not identified by the authors; the most probable species is shown
C:Comment: This protein is an endothelial-specific receptor and binds vascular endothel
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase Homolo
C:Keywords: ATP; embryo; growth factor receptor; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-138/Product: vascular endothelial growth factor receptor 1 #status predicted <MAR>
F:756-777/Domain: transmembrane #status predicted <TM>
F:823-1160/Domain: protein kinase homology <KIN>
F:831-839/Region: protein kinase ATP-binding motif

Query Match      7.8%; Score 102.5; DB 2; Length 1348;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 49; Conservative 33; Mismatches 73; Indels 59; Gaps 12;

QY      9 HLLVLTQALLPAAQTGNKVVLGKGDVLTCTASQKSIQFMW-KNSNOIKILNGQS 67
Db      536 HVTGRLINLQPRSQLTEK-----DNTSLQCTADKFTFKSLWYKLSHTV-----SQTP 584
QY      68 FLTGTG--PSKINDAADSRSLL-----WDQGFPLIKIKLKEDSDTYICEVDQKEE 117
Db      585 F--GGLEMPVCKNLDAQKLNATVSNVNGENVTLLEILIRISLQDGDYVCAIQDKKAK 641
QY      118 VQ-LLVGLTANSDTH-----LQGQSITLTLESPPGSSPSVQCRSPRG-----KN 162

```

```

Db      642 TQHCIVKHLTVQEPILPRVLGNLEQNTNI-----GERIEVLG-TVNGVPPPNITWPKN 694
QY      163 IQ-----GKTLVSQLELODSGTWTC 184
Db      695 SETLPEDSGIVLKDGKNTLTIRRVKEDGLGYTC 728

RESULT 78
WMM5R1
Biliary glycoprotein A precursor - mouse
N:Alternate names: carcinoembryonic antigen mmCGM1a; murine hepatitis virus receptor
C:Species: Mus musculus (house mouse)
C>Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text_change 22-Jun-1999
C:Accession: J01505; A49006; A41563; S11626; S11625
R:McCaig, K.; Rosenberg, M.; Nedellec, P.; Turbide, C.; Beauchemin, N.
Gene 127, 173-183, 1993
A:Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopro
A:Reference number: J01505; MUID:93273228; PMID:8500759
A:Accession: J01505
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-458 <MC>
R:McCaig, K.; Turbide, C.; Beauchemin, N.
Cell Growth Differ. 3, 165-174, 1992
A:Title: mmCGM1a: a mouse carcinoembryonic antigen gene family member, generated by alt
A:Reference number: A49006; MUID:92338096; PMID:1633107
A:Accession: A49006
A:Molecule type: mRNA
A:Residues: 1-458 <MC2>
A:Cross-references: GB:X15351; NID:G53020; PIDN:CAA33409.1; PID:G53021
A:Note: sequence extracted from NCBI backbone (NCBIN:109445, NCBIPI:109446)
R:Dvorkler, G.S.; Pensiero, M.N.; Cardelliho, C.B.; Williams, R.K.; Jiang, G.S.; Holm
U. Virol. 65, 6881-6891, 1991
A:Title: Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and h
A:Reference number: A41563; MUID:92046352; PMID:1719235
A:Accession: A41563
A:Molecule type: mRNA
A:Residues: 1-458 <DVE>
A:Cross-references: GB:M77196; NID:G194002; PIDN:AA37858.1; PID:G194003
R:Beauchemin, N.
submitted to the EMBL Data Library, May 1990
A:Reference number: S11626
A:Accession: S11626
A:Molecule type: mRNA
A:Residues: 1-82-458 <BEA>
A:Cross-references: EMBL:X15351
R:Beauchemin, N.; Turbide, C.; Afar, D.; Bell, J.; Raymond, M.; Stanners, C.P.; Fuks, A
Cancer Res. 49, 2017-2021, 1989
A:Title: A mouse analogue of the human carcinoembryonic antigen.
A:Reference number: S11625; MUID:89195121; PMID:2702644
A:Accession: S11625
A:Molecule type: mRNA
A:Residues: 1-82-321 <BE2>
A:Cross-references: EMBL:X15351
C:Comment: This protein is expressed at the cell surface and plays a determinant role in
C:Genetics:
A:Gene: Bgpa
C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C:Keywords: glycoprotein; receptor; transmembrane protein
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-458/Product: biliary glycoprotein A #status predicted <MHR>
F:160-219/Domain: immunoglobulin homology <IMM1>
F:254-303/Domain: immunoglobulin homology <IMM2>
F:339-396/Domain: immunoglobulin homology <IMM3>
F:429-447/Domain: transmembrane #status predicted <TMN>
F:449-558/Domain: intracellular #status predicted <CT>
F:71,89,104,148,152,199,206,210,226,258,290,294,304,317,333,375/Binding site: carbohydr
F:167-217,261-301,346-394/Disulfide bonds: #status predicted

Query Match      7.7%; Score 102; DB 1; Length 458;
Best Local Similarity 22.5%; Pred. No. 4;
Matches 42; Conservative 35; Mismatches 68; Indels 42; Gaps 10;

```

```

Oy 16 LALLPAATQGNKVVILGKKGDVTELTCTA-SOKKSIQFHMKNSNQIKILGNGSFLTKGPS 74
Db 143 ILKRNKNTSNNSNPV-EGDDSVSLTCDSTDPDNTNVMYSRN-----GESLSGSD- 191
Oy 75 KLNDRAUSRRSLMDQGNFPLIIRKLIKEDSDTYICEVE-----DOKEEVQV-LVEFG---- 124
Db 192 -----RLKLSBGNNTLTLNLTANDTGPYCETNPVSVNRSDPFLNITIIYGPPTP 242
Oy 125 LTANSDFHLLOGQSITLTLESPPGSSPSVQC-----RSRKNKIQQGKTLSSVQLQLQDS 179
Db 243 IISPSDIYHFGSNLNSCHA--ASNPPAQYFWLINEKPHASS-----QELFIPMITTNNS 296

Oy 180 GTWTCCTV 186
Db 297 GTYTCFV 303

RESULT 79
523969
cell adhesion molecule short form (cell-CAM105) - rat
N/Alternate names: C-CAM protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 22-Nov-1993 #sequence revision 01-Sep-1995 #text_change 23-Jul-1999
C/Accession: S23969; S32483; S38826; S10563; S32102
R/Cultic: O.; Huang, O.H.; Flanagan, D.; Hixson, D.; Lin, S.H.
Biochem. J. 285, 47-53, 1992
A/Title: Molecular cloning and expression of a new rat liver cell-CAM105 isoform. Differ
A/Reference number: S23969; MUID:92344597; PMID:1637321
A/Accession: S23969
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-458 <CU>
A/Cross-references: EMBL:Z12019; NID:955858; PIDN:CAA76054.1; PID:955859
R/Edlund, M.; Gaardsvoll, H.; Bock, E.; Oebdrink, B.
Eur. J. Biochem. 213, 1109-1116, 1993
A/Title: Different isoforms and stock-specific variants of the cell adhesion molecule C-
A/Reference number: S32483; MUID:93279310; PMID:8504806
A/Accession: S32483
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-458 <ED>
A/Cross-references: EMBL:X71122; NID:9287819; PIDN:CAA50435.1; PID:9287820
R/Cheung, P.H.; Culic, O.; Otu, Y.; Bailey, K.; Thompson, N.; Hixson, D.C.; Lin, S.H.
Biochem. J. 293, 427-435, 1993
A/Title: The cytoplasmic domain of C-CAM is required for C-CAM-mediated adhesion function
A/Reference number: S38826; MUID:94058980; PMID:8240240
A/Accession: S38826
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-48, 'A', 50-54, 'A', 56-69, 'G', 71-72, 'LNPV', 77-85, 'D', 87, 'M', 89, 'K', 91, 'G', 93-
A/Cross-references: EMBL:M92848; NID:920366; PIDN:AAA16783.1; PID:920367
R/Autivillius, M.; Hansen, O.C.; Lazrek, M.B.S.; Bock, E.; Oebdrink, B.
FEBS Lett. 264, 267-269, 1990
A/Title: The cell adhesion molecule cell-CAM 105 is an ecto-ATPase and a member of the
A/Reference number: S10563; MUID:90292222; PMID:2141577
A/Accession: S10563
A/Molecule type: protein
A/Residues: 'X', 58-66, 'A', 68, 121-124, 'F', 126, 'Q', 128-134, 'X', 136-138, 'X', 1356-160, 'X', 1362
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEA>
F:160-217/Domain: immunoglobulin homology <IMM1>
F:252-301/Domain: immunoglobulin homology <IMM2>
F:337-394/Domain: immunoglobulin homology <IMM3>

Query Match 7.7%; Score 102; DB 2; Length 458;
Best Local Similarity 21.0%; Pred. No. 4.8;
Matches 49; Conservative 33; Mismatches 85; Indels 66; Gaps 7;

Oy 33 KGDVTELTCTASQKKSIGFHMKNSNQIKILGNGSFLTKGPSKLNDRAUSRRSLMDQGNF 92
Db 251 QGSNLTNSCHADSNPPAQYFWLINEKLTSSQE----- 283


```

```

Oy 93 PLIKNKIEDSDTYICEVEDEQKEVQLVFGLTAN-----SDTHLL 134
Db 284 -LFISNTITNNSGYACFVNN-----TYTGLSRTTKVKNITVEPYTQSIQTNTTVK 335
Oy 135 QGOSLTTLTLESPPG-----SSPVQCRSPRKNKIQQGKTLSSVQLQLQDSGTWTCCTV 186
Db 336 ELGSVTLTLCRSKOTGVSVRWLFNSQSLQ--TDRWTLISQDNSTLRIIDPKREDADYQCEI 394
Oy 187 LQNKQVEFKIDIVPRASALPAPPTG-SALPDQPTASALPDPPASALPAPALA 238
Db 395 ---SNPVSRISHPIKIDVIPDPQTGNSGLSEGAIAGIVIGSVAGVALLA 444

RESULT 80
568177
C-CAM2a protein isoform precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Feb-1997 #sequence revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: S68177
R/Lucka, L.; Cichocka, I.; Baeumler, K.; Bechler, K.; Reutter, W.
Eur. J. Biochem. 234, 527-535, 1995
A/Title: A short isoform of carcinoembryonic-antigen-related rat liver cell-cell adhesi
A/Reference number: S68177; MUID:96128184; PMID:8536699
A/Accession: S68177
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-458 <LU>
A/Cross-references: EMBL:X91137; NID:9160272; PIDN:CAA62577.1; PID:91160273
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEA>
F:133/Domain: signal sequence #status predicted <SIG>
F:34-458/Product: C-CAM2a protein isoform #status predicted <MAT>
F:252-301/Domain: immunoglobulin homology <IMM>

Query Match 7.7%; Score 102; DB 2; Length 458;
Best Local Similarity 21.0%; Pred. No. 4.8;
Matches 49; Conservative 33; Mismatches 85; Indels 66; Gaps 7;

Oy 33 KGDVTELTCTASQKKSIGFHMKNSNQIKILGNGSFLTKGPSKLNDRAUSRRSLMDQGNF 92
Db 251 QGSNLTNSCHADSNPPAQYFWLINEKLTSSQE----- 283
Oy 93 PLIKNKIEDSDTYICEVEDEQKEVQLVFGLTAN-----SDTHLL 134
Db 284 -LFISNTITNNSGYACFVNN-----TYTGLSRTTKVKNITVEPYTQSIQTNTTVK 335
Oy 135 QGOSLTTLTLESPPG-----SSPVQCRSPRKNKIQQGKTLSSVQLQLQDSGTWTCCTV 186
Db 336 ELGSVTLTLCRSKOTGVSVRWLFNSQSLQ--TDRWTLISQDNSTLRIIDPKREDADYQCEI 394
Oy 187 LQNKQVEFKIDIVPRASALPAPPTG-SALPDQPTASALPDPPASALPAPALA 238
Db 395 ---SNPVSRISHPIKIDVIPDPQTGNSGLSEGAIAGIVIGSVAGVALLA 444

RESULT 81
JC1508
biliary glycoprotein D - mouse
N/Alternate names: biliary glycoprotein 1
C/Species: Mus musculus (house mouse)
C/Date: 24-Feb-1994 #sequence revision 24-Feb-1994 #text_change 23-Jul-1999
C/Accession: JC1508; S65940; S36852
R/McCaig, K.; Rosenberg, W.; Nedellec, P.; Turbide, C.; Beauchemin, N.
Gene 127, 173-183, 1993
A/Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopr
A/Reference number: JC1505; MUID:93273228; PMID:8500759
A/Accession: JC1508
A/Molecule type: mRNA
A/Residues: 1-521 <MC>
A/Cross-references: EMBL:X67279; NID:950170; PIDN:CAA7666.1; PID:950171
A/Experimental source: strain CD1, tissue colon
R/Nedellec, P.; Turbide, C.; Beauchemin, N.
Eur. J. Biochem. 231, 104-114, 1995


```

A:Title: Characterization and transcriptional activity of the mouse biliary glycoprotein  
A:Reference number: S65939; MUID:95354676; PMID:7628460  
A:Accession: S65940  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-21 <NED>  
A:Cross-references: EMBL:X84054; NID:g1039337  
A:Experimental source: strain BALB/c  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1992  
A:Note: only a part of the coding sequence is given  
C:Comment: This protein is expressed at the cell surface and plays a determinant role in  
C:Genetics:  
A:Gene: Bgpd; bgp1  
A:Map position: 7  
C:Keywords: glycoprotein; receptor; transmembrane protein  
C:Keywords: glycoprotein; receptor; transmembrane protein  
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>  
F:160-219/Domain: immunoglobulin homology <IWM1>  
F:254-303/Domain: immunoglobulin homology <IWM2>  
F:339-396/Domain: immunoglobulin homology <IWM3>  
F:71,89,104,148,199,206,210,226,258,290,294,304,317,333,375/Binding site: carbohydrate

Query Match 7.7%; Score 102; DB 2; Length 521;  
Best local similarity 22.5%; Pred. No. 5.5;  
Matches 42; Conservative 35; Mismatches 68; Indels 42; Gaps 10;

16 LALLPAALQGNKVVLGKKGDVIELTCTA-SQKSIQPHMKNSNQIKLQGSFLTGPS 74  
143 ILKPNITSNNSNPV-EGSDSVSLTCDSTYDPDNIINVLMSN-----GSLSGD- 191  
75 KLNDRAISRRLMQGNFPLIKLKIEDSTYICEV-----DQKEVOL-LVFG--- 124  
192 -----RLKLSGKNTLLNLVTRNDTGPVCEFRNPVSVRSPPFSLNIIYGPDP 242  
125 LTANSDFLLQGSILTLTLSPGSSPSVQC-----RSPRKNIGGKTLVSOLFQDS 179  
243 IISSDIYLRHGSNLTLSCHA--ASNPPAQVFWLINKEPHASS---QELFPIRITNNS 256  
180 GTWTCTV 186  
297 GTYTCEV 303

RESULT 82  
18346  
Elastic titin - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
A:Accession: I38346  
R:Label: S.; Kolmerer, B.  
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
A:Reference number: A57430; MUID:96026330; PMID:7569978  
A:Accession: I38346  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-7962 <RES>  
A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427  
C:Genetics:  
A:Gene: GDB:TTN  
A:Cross-references: GDB:127867; OMIM:188840  
A:Map position: 2q31-2q31

Query Match 7.7%; Score 102; DB 2; Length 7962;  
Best local similarity 23.6%; Pred. No. 1.4e+02;  
Matches 59; Conservative 22; Mismatches 81; Indels 88; Gaps 11;

11 LVLVQLALPAPATQGNKVVLGKGDVIELTCTASQKSIQPH-----KNSNQIKLG 63  
Db 4245 LSVLEPATIVKPSIKVT---TGDTCTLECTVAGTPELSTKWKFGDKELTSDNKYKI-- 4299  
QY 64 NQGSFLTGPSKLNDRAISRRLMQGNFPLIKLKIEDSTYICEVEDQKEVOLVF 123

```

Db      4300 ---SFFKVSQ-----LKIINAPSDSGVSEFVONP----- 4328
Qy      124 GLTANSDTHLQGSLTLTLESP-----GSSPSVQCR---SP-----R 159
Db      4329 -VKDSDCTASLQVSDRTV---PSPFTRKLKETNGSLGSSVVMCKCYGSPPLSVSEFHE 4383
Qy      160 GKNIQGGK-----TLVSQLELQDSGTGTCTVLQONOKVEFKFIDIVPRASALPA 208
Db      4384 GNEISSGRKXQKTLTLNTQCALTVNMLESDSDSGDYCTIATNMAGSDSCAPLTVR-----E 4438
Qy      209 PPTGSALPDP 218
Db      4439 PPSFVQKPPD 4448

RESULT 83
P1066
Ig light chain V region (clone 165.14) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_rev1stion 30-Sep-1993 #text_change 21-Jan-2000
C:Accession: P1066
R:Titman, D.M.; Jou, N.T.; Hall, R.J.; Marlon, T.N.
J: Exp. Med. 176: 761-779, 1992
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B
A:Reference number: PH0571; MUID:92381444; PMID:1512540
A:Accession: P1066
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Reidues: 1-98 <TIL>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match      7.7%; Score 101.5; DB 2; Length 98;
Best Local Similarity 31.8%; Pred. No. 0.83;
Matches 35; Conservative 7; Mismatches 39; Indels 29; Gaps 3;

Qy      34 GDTVELTCTASQSKSIQFMKNSNQIKILGNQGSFLTKGPKSLNDRADSRSLMDQG-NF 92
Db      16 GDTITTCRASQININIMLWYQQR-----GNIPKLLIYKASNLHTGVPSRFSGSGSTDP 71

Qy      93 PLTIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGSLLTLT 142
Db      72 TLTISSQLQPEDIATYYC-----LQGQSYPLT 77

RESULT 84
IUMSNG
neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
N:Alternate.names: NCAM-120
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000
C:Accession: A29673; S00382; A44290
R:Barthels, D.; Santoni, M.J.; Wille, W.; Rupprecht, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
R:Barthels, D.; Santoni, M.J.; Wille, W.; Rupprecht, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
EMBO J. 7: 625-632, 1988
A:Title: Differential splicing and alternative polyadenylation generates distinct NCAM
A:Reference number: S00382; MUID:88283628; PMID:3396534
A:Accession: S00382
A:Molecule type: DNA
A:Reidues: 642-656, 'D', 658-725 <DA2>
A:Cross-references: EMBL:X07195
R:Rougon, G.; Marchak, D.R.
J: Biol. Chem. 261: 3396-3401, 1986
A:Title: Structural and immunological characterization of the amino-terminal domain of

```

A:Reference number: A44290; MUID:86140120; PMID:3512556  
A:Accession: A44290  
A:Molecule type: protein  
A:Residues: 20-36 <R0U>  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol  
C:Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:JUMS  
C:Genetics:  
A:Gene: NCAM  
A:Map position: 9  
A:Interons: 701/1  
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:34-98/Domain: immunoglobulin homology <IM1>  
F:132-191/Domain: immunoglobulin homology <IM2>  
F:152-156/Region: heparin binding #status predicted  
F:161-165/Region: heparin binding #status predicted  
F:228-290/Domain: immunoglobulin homology <IMW3>  
F:263-272/Region: NCAM binding #status predicted  
F:323-388/Domain: immunoglobulin homology <IMW4>  
F:420-482/Domain: immunoglobulin homology <IMW5>  
F:519-556/Domain: fibronectin type III repeat homology <FN3A>  
F:625-665/Domain: fibronectin type III repeat homology <FN3B>  
F:41-96,139-189,235-288,330-386,427-480/Dissulfide bonds: #status predicted  
F:222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.7%; Score 101.5; DB 1; Length 725;  
Best Local Similarity 19.9%; Pred. No. 8.9;  
Matches 46; Conservative 42; Mismatches 94; Indels 49; Gaps 10;

3 RG-VPFRLHLVLQALLPATQGNKVLGKSGDTVELTCTASQKSIQPHW-KNSNQIK 60  
196 RGEINFDIYIVNPPVQARQSIYATNANLGQSVTLVCDAGPPEPTMSWTGDGP 255  
Db 1 LGNGSFLTKGPEKNDRAD--SRSLMDQGNFPLIINKLKIEDSDTYICEV---DQ 114  
256 -----NEEDERSRSVSDSE--VTIRNVKNDKDEAYVCIANKKGEQ 297  
Qy 115 KEVQLVFG---LTANSDFHLQGSLLTLESPPSSPSVOCR-----SPGKNIQ 164  
Db 298 DASHLKFAKPKITTYENQTMELBEQVTLTCEASGDPPIPTWTRNISSEEDLD 357  
Qy 165 GG-----KTLVSQLELDQSGTWC---TVLQNKVKEFKIDIVR 202  
Db 358 GHMVRSHARVSLTKSIQYRDAGEYWCASNTTIGDSQSIDLEFOYAPK 408

RESULT 85  
JURTNC  
neural cell adhesion molecule short domain form precursor - rat  
N/Alternate names: NCAM-140  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 22-Jun-1999  
C/Accession: S00846; B37795; I58136  
R/Small: S.J.; Shull, G.E.; Santoni, M.J.; Akeson, R.  
J. Cell Biol. 105, 2335-2345, 1987  
A:Title: Identification of a cDNA clone that contains the complete coding sequence for a  
A:Reference number: S00846; MUID:86059265; PMID:3680385  
A:Accession: S00846  
A:Molecule type: mRNA  
A:Residues: 1-858 <SMA>  
A:Cross-references: EMBL:X06564  
R/Small: S.J.; Akeson, R.  
J. Cell Biol. 111, 2089-2096, 1990  
A:Title: Expression of the unique NCAM VASE exon is independently regulated in distinct  
A:Reference number: A37795; MUID:91035620; PMID:1699951  
A:Accession: B37795  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 340-381 <SM2>  
R/Small: S.J.; Haines, S.L.; Akeson, R.A.  
Neuron 1, 1007-1017, 1988  
A:Title: Polypeptide variation in an N-CAM extracellular immunoglobulin-like fold is de

A:Reference number: I58136; MUID:90166485; PMID:2483093  
A:Accession: I58136  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 355-364 <RES>  
A:Cross-references: GB:M32611; NID:g205643; PIDN:AAA41679.1; PID:g205644  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mo  
C:Comment: Various forms of NCAM are produced by alternative splicing.  
C:Genetics:  
A:Gene: NCAM  
A:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; imm  
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; s  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-858/Product: neural cell adhesion molecule, short domain form #status predicted <M  
F:34-98/Domain: immunoglobulin homology <IM1>  
F:132-191/Domain: immunoglobulin homology <IM2>  
F:152-156/Region: heparin binding #status predicted  
F:161-165/Region: heparin binding #status predicted  
F:228-290/Domain: immunoglobulin homology <IMW3>  
F:263-272/Region: NCAM binding #status predicted  
F:323-398/Domain: immunoglobulin homology <IMW4>  
F:430-492/Domain: immunoglobulin homology <IMW5>  
F:529-606/Domain: fibronectin type III repeat homology <FN3A>  
F:635-695/Domain: fibronectin type III repeat homology <FN3B>  
F:722-739/Domain: transmembrane #status predicted <TM>  
F:740-858/Domain: intracellular #status predicted <INT>  
F:41-96,139-189,235-288,330-396,437-490/Dissulfide bonds: #status predicted  
F:222,316,348,434,460,489/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.7%; Score 101.5; DB 1; Length 858;  
Best Local Similarity 18.0%; Pred. No. 11;  
Matches 43; Conservative 42; Mismatches 99; Indels 55; Gaps 8;

3 RG-VPFRLHLVLQALLPATQGNKVLGKSGDTVELTCTASQKSIQPHW-KNSNQIK 60  
196 RGEINFDIYIVNPPVQARQSIYATNANLGQSVTLVCDAGPPEPTMSWTGDGP 255  
Db 1 LGNGSFLTKGPEKNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEV---DQKE 116  
256 -----NEEDDEKHIFSDSSSELTIRNVKNDKDEAYVCIANKKGEQ 299  
Qy 117 EVQLVFG---LTANSDFHLQGSLLTLESPPSSPSVOCR-----SPGKNIQ 157  
Db 300 SIHLKFAKPKITTYENQTMELBEQVTLTCEASGDPPIPTWTRNISSEKASWTR 359  
Qy 158 PRGKNIOG-----KTLVSQLELDQSGTWC---TVLQNKVKEFKIDIVR 202  
Db 360 PERKETLDGHMVRSHARVSLTKSIQYRDAGEYWCASNTTIGDSQSVYLEVQYAPK 418

RESULT 86  
JUMSNL  
neural cell adhesion molecule 1 precursor, long domain splice form - mouse  
N/Alternate names: NCAM-180  
C/Species: Mus musculus (house mouse)  
C/Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text\_change 31-Dec-2000  
C/Accession: A29673; S00844; S00384; A28281; A44290; S00383  
R/Bartels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaux, J.C.; Hirsch, M.R.; Fontes  
EMBO J. 6, 907-914, 1987  
A:Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,00  
A:Reference number: A29673; MUID:87246524; PMID:3595563  
A:Accession: A29673  
A:Molecule type: mRNA  
A:Residues: 1-548, 'T', 550-571, 'T', 573-574, 'D', 576-588, 'WQPS', 593, 'S', 595-599, 'P', 601, 'L  
A:Cross-references: EMBL:Y00051; NID:g53342; PIDN:CAA68263.1; PID:g53343  
R/Santoni, M.J.; Bartels, D.; Bartel, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.;  
Nucleic Acids Res. 15, 8621-8641, 1987  
A:Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse neu  
A:Reference number: S00844; MUID:88067687; PMID:3684567  
A:Accession: S00844  
A:Molecule type: mRNA

A:Residues: 529-809,1077-1115 <SAN>  
 A:Cross-references: EMBL:X06328; NID:G53322; PIDN:CAA29641.1; PID:G817984  
 R:Barbas, J.A.; Chaux, J.C.; Steinmetz, M.; Gortidis, C.  
 EMBL J. 7, 625-632, 1988  
 A:Title: Differential splicing and alternative polyadenylation generates distinct NCAM  
 A:Reference number: S00382; MUID:88263628; PMID:3396534  
 A:Accession: S00384  
 A:Molecule type: DNA  
 A:Residues: 642-1115 <BAR>  
 A:Cross-references: EMBL:X07195  
 R:Barthels, D.; Vopper, G.; Wille, W.  
 Nucleic Acids Res. 16, 4217-4225, 1988  
 A:Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mouse.  
 A:Reference number: A28281; MUID:88247737; PMID:2454455  
 A:Accession: A28281  
 A:Molecule type: mRNA  
 A:Residues: 804-1081 <BA3>  
 A:Cross-references: EMBL:X07244; NID:G53321; PIDN:CAA30230.1; PID:G929720  
 R:Rougon, G.; Marshak, D.R.  
 J. Biol. Chem. 261, 3395-3401, 1986  
 A:Title: Structural and immunological characterization of the amino-terminal domain of  
 A:Reference number: A44290; MUID:86140120; PMID:3512556  
 A:Accession: A44290  
 A:Molecule type: protein  
 A:Residues: 20-36 <ROU>  
 C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol  
 C:Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IMM  
 C:Genetics:  
 A:Gene: NCAM  
 A:Map position: 9  
 A:Intons: 643/3; 701/1; 770/2; 809/2; 1076/2  
 C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
 C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; st  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1115/Product: neural cell adhesion molecule, long domain splice form #status experi  
 F:20-809,1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta  
 F:20-711/Domain: extracellular #status predicted <EXT>  
 F:34-98/Domain: immunoglobulin homology <IMM1>  
 F:132-191/Domain: immunoglobulin homology <IMM2>  
 F:152-156/Region: heparin binding #status predicted  
 F:161-165/Region: heparin binding #status predicted  
 F:228-290/Domain: immunoglobulin homology <IMM3>  
 F:262-272/Region: NCAM binding #status predicted  
 F:333-388/Domain: immunoglobulin homology <IMM4>  
 F:420-482/Domain: immunoglobulin homology <IMM5>  
 F:519-596/Domain: fibronectin type III repeat homology <FN3A>  
 F:625-685/Domain: fibronectin type III repeat homology <FN3B>  
 F:712-729/Domain: transmembrane #status predicted <TM>  
 F:720-1115/Domain: intracellular #status predicted <INT>  
 F:41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted  
 F:222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.7%; Score 101.5; DB 1; Length 1115;  
 Best Local Similarity 19.9%; Pred. No. 15;  
 Matches 46; Conservative 42; Mismatches 94; Indels 49; Gaps 10;

QY 3 RG-VPRPRLVQLALPAPATGKAVVVGKGGDTVELTCTASQKSIQFHW-KNSNQIK 60  
 DB 196 RGEINFDIOIVIVVWPPTVAPROSIVATNATLQGSVTLVADAGFPPTMSWTKDGP 255  
 QY 61 ILGNQGSFLTKGPKSLNDRAD--SRSLMDQGNPPLIKLKIEDSDTYICEVE---DQ 114  
 DB 256 -----NEEDERSRSSVDSSE--VIRIVDKNDDEYVICENKAKGEQ 257  
 QY 115 KEEVOLLVFG---LTNSDTHLLQGSFLTLESPPGSSPVQCR-----SPRKNIO 164  
 DB 298 DASHLKVFAPKPKITTYENQAMLEEQVTLTCBASDPIPSITWRSTNINSEEGDLD 357  
 QY 165 GG-----KTLVSQLELDGSGTTC---TVLQNKQKVEKIDIVR 202  
 DB 358 GHNVVSRHARVSSLTKSIQYRDAGEYWCASNTIGDSIDLEFQYAK 408

RESULT 87  
 S10004  
 hypothetical protein 6 - fowl adenovirus 1  
 C:Species: Aviatadenovirus gall (fowl adenovirus 1, CEL0)  
 C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 08-Oct-1999  
 C:Accession: S10004  
 R:Akopian, T.A.; Kruglyak, V.A.; Rikina, M.B.; Naroditsky, B.S.; Tikhonenko, T.I.  
 Nucleic Acids Res. 18, 2825, 1990  
 A:Title: Sequence of an avian adenovirus (CEL0) DNA fragment (0-11.2%).  
 A:Reference number: S10004; MUID:90251474; PMID:2160072  
 A:Accession: S10004  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-135 <AKO>  
 A:Cross-references: EMBL:X17217; NID:G58537; PIDN:CAA35086.1; PID:G58538

Query Match 7.7%; Score 101; DB 2; Length 135;  
 Best Local Similarity 27.5%; Pred. No. 13;  
 Matches 38; Conservative 21; Mismatches 65; Indels 14; Gaps 4;

QY 11 LVVLQALIPAT--QGNKVLGKGGDTVELTCTASQKSIQFHWKNSNQIKLGNQSF 68  
 DB 1 MLTLVLLVGYTLADHPTLVAPKGSITELGVAKQYFRFNG-LKIVIAEMSS 59  
 QY 69 LTKGPKSLNDRADSRSLMDQGNPPLIKLKIEDSDTYICEVEQKEVOLLVFGLTAN 128  
 DB 60 TQLEIKFPDNGFONRSENPPTKHLTLTNASYESDGTSLHOENDG-----TEH 110

QY 129 SDTH--LLQGSFLTLE 144  
 DB 111 TDFKVIQVQMSLYTLQ 128

RESULT 88  
 S22383  
 axonin 1 precursor - chicken  
 N:Alternate names: neural cell adhesion molecule AxCAM  
 C:Species: Gallus gallus (chicken)  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000  
 C:Accession: S22383; S34107; S69332; S22128  
 R:Zuellig, R.A.; Rader, C.; Schroeder, A.; Kalousek, M.B.; von Bohnen und Halbach, F.;  
 Eur. J. Biochem. 204, 453-463, 1992  
 A:Title: The axonally secreted cell adhesion molecule, axonin-1. Primary structure, immu  
 A:Reference number: S22383; MUID:92174898; PMID:1311675  
 A:Accession: S22383  
 A:Molecule type: mRNA  
 A:Residues: 1-1036 <ZUE1>  
 A:Cross-references: EMBL:X63101; NID:G62852; PIDN:CAA44815.1; PID:G62853  
 A:Accession: S34107  
 A:Molecule type: protein  
 A:Residues: 29-49,51-80;84-95;100-117,120-128,130-141,143-176,243-254,256-296,303-336,337  
 R:Giger, R.U.; Vogt, L.; Zuellig, R.A.; Rader, C.; Henenhan-Beatty, A.; Wolter, D.P.; Sch  
 Eur. J. Biochem. 227, 617-628, 1995  
 A:Title: The gene of chicken axonin-1. Complete structure and analysis of the promoter.  
 A:Reference number: S69332; MUID:95172044; PMID:7867620  
 A:Accession: S69332  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1002-1036 <GIG>  
 A:Cross-references: EMBL:X79607  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994  
 C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology  
 C:Keywords: cell adhesion  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-1036/Product: axonin 1 #status predicted <MAT>  
 F:336-392/Domain: immunoglobulin homology <IMM>

Query Match 7.7%; Score 101; DB 2; Length 1036;  
 Best Local Similarity 23.2%; Pred. No. 15;  
 Matches 53; Conservative 30; Mismatches 87; Indels 58; Gaps 9;

QY 34 GDTVELTCTASQKSIQFHWKNSNQIKLGNQGSFLTKGPKSLNDRADSRSLMDQGNFP 93



Db 249 GQMTLLECAFGFNPPQJIKMR-----KLDGQ-----ISKMLSEPL 285  
 Qy 94 LIINKLKIEDSDTYICEVEDOK---EEVQLLVFG---LTANSPTHLIQGSLTFTES 145  
 Db 286 LHIQVDEDEDEGTCECAENIKGRDTYQGRITIIHQPMDLVITTEADIGSGLHMSCVA 345  
 Qy 146 PPGSSPSVQ-----CRSPRGKNIQGGKITLSVSOLELSDSTMTCTYVLQONOKV----- 193  
 Db 346 SGKRPAPRWLRDQGPLASQNRIEVSGGE-LRFSKVLVEDSSGMYOCVAENKKGITVYASAE 404  
 Qy 194 -----EPKIDIVRASALPAPPTGSALPPQRTASALPDPAPASAL 233  
 Db 405 LTVQALADPFRLLNPVKR--LTPAASGKVIITPCQPPAA---PKATVL 446

RESULT 89

T19821  
hypothetical protein ZC101.2e - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 23-Sep-2002  
A/Accession: T19821; T19819; T19820; T27490; T27488; T27489; T27487; A47648; B47648; C47648; D47648; E47648; F47648; G47648; H47648; I47648; J47648; K47648; L47648; M47648; N47648; O47648; P47648; Q47648; R47648; S47648; T47648; U47648; V47648; W47648; X47648; Y47648; Z47648; AA47648; AB47648; AC47648; AD47648; AE47648; AF47648; AG47648; AH47648; AI47648; AJ47648; AK47648; AL47648; AM47648; AN47648; AO47648; AP47648; AQ47648; AR47648; AS47648; AT47648; AU47648; AV47648; AW47648; AX47648; AY47648; AZ47648; BA47648; BB47648; BC47648; BD47648; BE47648; BF47648; BG47648; BH47648; BI47648; BJ47648; BK47648; BL47648; BM47648; BN47648; BO47648; BP47648; BQ47648; BR47648; BS47648; BT47648; BU47648; BV47648; BW47648; BX47648; BY47648; BZ47648; CA47648; CB47648; CC47648; CD47648; CE47648; CF47648; CG47648; CH47648; CI47648; CJ47648; CK47648; CL47648; CM47648; CN47648; CO47648; CP47648; CQ47648; CR47648; CS47648; CT47648; CU47648; CV47648; CW47648; CX47648; CY47648; CZ47648; DA47648; DB47648; DC47648; DD47648; DE47648; DF47648; DG47648; DH47648; DI47648; DJ47648; DK47648; DL47648; DM47648; DN47648; DO47648; DP47648; DQ47648; DR47648; DS47648; DT47648; DU47648; DV47648; DW47648; DX47648; DY47648; DZ47648; EA47648; EB47648; EC47648; ED47648; EE47648; EF47648; EG47648; EH47648; EI47648; EJ47648; EK47648; EL47648; EM47648; EN47648; EO47648; EP47648; EQ47648; ER47648; ES47648; ET47648; EU47648; EV47648; EW47648; EX47648; EY47648; EZ47648; FA47648; FB47648; FC47648; FD47648; FE47648; FF47648; FG47648; FH47648; FI47648; FJ47648; FK47648; FL47648; FM47648; FN47648; FO47648; FP47648; FQ47648; FR47648; FS47648; FT47648; FU47648; FV47648; FW47648; FX47648; FY47648; FZ47648; GA47648; GB47648; GC47648; GD47648; GE47648; GF47648; GG47648; GH47648; GI47648; GJ47648; GK47648; GL47648; GM47648; GN47648; GO47648; GP47648; GQ47648; GR47648; GS47648; GT47648; GU47648; GV47648; GW47648; GX47648; GY47648; GZ47648; HA47648; HB47648; HC47648; HD47648; HE47648; HF47648; HG47648; HH47648; HI47648; HJ47648; HK47648; HL47648; HM47648; HN47648; HO47648; HP47648; HQ47648; HR47648; HS47648; HT47648; HU47648; HV47648; HW47648; HX47648; HY47648; HZ47648; IA47648; IB47648; IC47648; ID47648; IE47648; IF47648; IG47648; IH47648; II47648; IJ47648; IK47648; IL47648; IM47648; IN47648; IO47648; IP47648; IQ47648; IR47648; IS47648; IT47648; IU47648; IV47648; IW47648; IX47648; IY47648; IZ47648; JA47648; JB47648; JC47648; JD47648; JE47648; JF47648; JG47648; JH47648; JI47648; JJ47648; JK47648; JL47648; JM47648; JN47648; JO47648; JP47648; JQ47648; JR47648; JS47648; JT47648; JU47648; JV47648; JW47648; JX47648; JY47648; JZ47648; KA47648; KB47648; KC47648; KD47648; KE47648; KF47648; KG47648; KH47648; KI47648; KJ47648; KK47648; KL47648; KM47648; KN47648; KO47648; KP47648; KQ47648; KR47648; KS47648; KT47648; KU47648; KV47648; KW47648; KX47648; KY47648; KZ47648; LA47648; LB47648; LC47648; LD47648; LE47648; LF47648; LG47648; LH47648; LI47648; LJ47648; LK47648; LL47648; LM47648; LN47648; LO47648; LP47648; LQ47648; LR47648; LS47648; LT47648; LU47648; LV47648; LW47648; LX47648; LY47648; LZ47648; MA47648; MB47648; MC47648; MD47648; ME47648; MF47648; MG47648; MH47648; MI47648; MJ47648; MK47648; ML47648; MM47648; MN47648; MO47648; MP47648; MQ47648; MR47648; MS47648; MT47648; MU47648; MV47648; MW47648; MX47648; MY47648; MZ47648; NA47648; NB47648; NC47648; ND47648; NE47648; NF47648; NG47648; NH47648; NI47648; NJ47648; NK47648; NL47648; NM47648; NO47648; NP47648; NQ47648; NR47648; NS47648; NT47648; NU47648; NV47648; NW47648; NX47648; NY47648; NZ47648; OA47648; OB47648; OC47648; OD47648; OE47648; OF47648; OG47648; OH47648; OI47648; OJ47648; OK47648; OL47648; OM47648; ON47648; OO47648; OP47648; OQ47648; OR47648; OS47648; OT47648; OU47648; OV47648; OW47648; OX47648; OY47648; OZ47648; PA47648; PB47648; PC47648; PD47648; PE47648; PF47648; PG47648; PH47648; PI47648; PJ47648; PK47648; PL47648; PM47648; PN47648; PO47648; PP47648; PQ47648; PR47648; PS47648; PT47648; PU47648; PV47648; PW47648; PX47648; PY47648; PZ47648; QA47648; QB47648; QC47648; QD47648; QE47648; QF47648; QG47648; QH47648; QI47648; QJ47648; QK47648; QL47648; QM47648; QN47648; QO47648; QP47648; QQ47648; QR47648; QS47648; QT47648; QU47648; QV47648; QW47648; QX47648; QY47648; QZ47648; RA47648; RB47648; RC47648; RD47648; RE47648; RF47648; RG47648; RH47648; RI47648; RJ47648; RK47648; RL47648; RM47648; RN47648; RO47648; RP47648; RQ47648; RR47648; RS47648; RT47648; RU47648; RV47648; RW47648; RX47648; RY47648; RZ47648; SA47648; SB47648; SC47648; SD47648; SE47648; SF47648; SG47648; SH47648; SI47648; SJ47648; SK47648; SL47648; SM47648; SN47648; SO47648; SP47648; SQ47648; SR47648; SS47648; ST47648; SU47648; SV47648; SW47648; SX47648; SY47648; SZ47648; TA47648; TB47648; TC47648; TD47648; TE47648; TF47648; TG47648; TH47648; TI47648; TJ47648; TK47648; TL47648; TM47648; TN47648; TO47648; TP47648; TQ47648; TR47648; TS47648; TT47648; TU47648; TV47648; TW47648; TX47648; TY47648; TZ47648; UA47648; UB47648; UC47648; UD47648; UE47648; UF47648; UG47648; UH47648; UI

A:Accession: A97648  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-546,'P',548-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRRHRRNAONGPLS'  
A:Cross-references: GB:L13458  
A:Accession: B47648  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-546,'P',548-2198,'D',2230,'NAR',2294,'L',2296,'WHAIE',2302-2303,'V',2305,  
1,'ANIV',2516-2517,'LOQ',2522,'IDG',2526,'S',2528,'SRGFHV',2535,'F'<RO2>  
A:Cross-references: GB:L13458  
A:Accession: C47648  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-546,'P',548-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545  
A:Cross-references: GB:L13458  
C:Genetics:  
A:Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b  
A:Map position: 2  
A:Introns: 32/1, 134/1, 225/1, 335/2, 450/3, 739/3, 830/3, 860/2, 1064/2, 1129/1, 1158/  
2, 2613/1, 2684/1, 2757/1, 2813/3, 2863/4, 2900/3, 3084/1, 3176/1, 3250/2  
C:Superfamily: LR11 protein; laminin-type EGF-like homology / LDL receptor ligand-bindin  
C:Keywords: extracellular matrix  
F:149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:199-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F:955-1002/Domain: laminin-type EGF-like homology <LEG1>  
F:1011-1058/Domain: laminin-type EGF-like homology <LEG2>

Query Match 7.7%; Score 101; DB 2; Length 3375;  
Best Local Similarity 18.6%; Pred. No. 60;  
Matches 50; Conservative 37; Mismatches 106; Indels 76; Gaps 10;

Dy 2 NRGVPFHLLVLDTALLPATGKVKVGKKGVETVELTASOKSKSIOPFMKNSNQIKI 61  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1581 NRPEPSNPARIYSPRIPIIDPAEQTV--PEGSPFKRCVPRGHSPSQLTRFRVS---- 1634  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy 62 LGNGSEFLTKGPSKLNDRAHSRSLSMDGNFPILLIKLKLTEDSDTYICEVD-----Q 114  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dd 1635 -----GQLNEDADENNGI-----LAVQAELTDEGDYICTARDPTGAPID 1675  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy 115 KEEOVLVFLGTANSSTHLLOGQSILTLES---PPSSPVQCSP----- 158  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dd 1676 STPATVTATAAAPQVEARPDPQHVIPTQTIPEDDPARIQCTVPGNPSAOHLSFER 1735  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy 159 -RGKNIGOGKT-----LSVSOLEDSGTWCITCLONOKVEFKDIIVPASALPAPTGG 212  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dd 1736 VDGKGLPFSGSSDDRGLTISTYTQIDAGGYCLX-----SPENSPEVK 1778  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dy 213 SALPDQTASALPD--PPAASALPALAV 239  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Dd 1779 T--NPSTLVNTPEGTPRPVATPPLLSV 1804  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

F:23-88/Disulfide bonds: #status predicted

Query Match 7.6%; Score 100.5; DB 1; Length 108;  
Best Local Similarity 39.0%; Pred. No. 1.1;  
Matches 30; Conservative 5; Mismatches 37; Indels 5; Gaps 2;

QY 34 GDFVETLCTASQKSIQIFHKNSNQIKILNQSGFLTKGSKLNDPDRSRSLMDQG-NP 92  
16 GDFVETLCTASQKSIQIFHKNSNQIKILNQSGFLTKGSKLNDPDRSRSLMDQG-NP 92

QY 93 PLIKNKLIKEDSDTYIC 109  
72 TLITISLEDEDMATYFC 88

RESULT 91

C42632  
cell adhesion molecule apCAM (clone d12) - California sea hare

C/Species: Aplysia californica (California sea hare)

C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-2000

C/Accession: C42632

R:Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel, E.R.  
Science 256, 638-644, 1992

A/Title: Modulation of an NCAM-related adhesion molecule with long-term synaptic plastic

A/Reference number: A42632; PMID:92263095; PMID:1585176

A/Accession: C42632

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-765 <MAY>

A/Experimental source: CNS

A/Note: sequence extracted from NCBI backbone (NCBIP:101351)

C/Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; imm

Query Match 7.6%; Score 100.5; DB 2; Length 765;  
Best Local Similarity 21.4%; Pred. No. 11;  
Matches 50; Conservative 34; Mismatches 79; Indels 71; Gaps 11;

QY 29 VLGKGDVETLCTASQKSIQIFHKNSNQIKILNQSGFLTKGSKLNDPDRSRSLMD 88  
136 ILGEGE--EVEGVSGKPAFTVTKFENNTKIEAGE--KYTALN----- 176

QY 89 QGNFPLIKNKLIKEDSDTYICEV-----EDQKEEVQLVFGI-----TANSDTHLQ 135  
177 ----KLIKDLSEDTKIKYLCIDIVDTGETKDFIDFTVVKPTALPTIHPDNPV- 231

QY 136 GQSLTLTLESPPSSPSVQCRS---PRKNIQSGKTLVSQELQDSGTWCTV----- 186  
232 GDEVKITCQATGVPPTVQFKGDMVWVTDENVNG-VLTINPLKTTDQATYTCIATNKG 290

QY 187 -LQNGKVEFKIDIVPRASALPAPT-----GSALPDPOFASALPDP 227  
291 FAESNTLTVK-----PTTEDEMYTAVSGLTITTTAKGDPEP 333

RESULT 92

A35963  
protein-tyrosine kinase (EC 2.7.1.112) cck2 precursor - chicken

C/Species: Gallus gallus (chicken)

C/Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 16-Jul-1999

C/Accession: A35963

R:Pasquale, E.B.

A/Title: A distinctive family of embryonic protein-tyrosine kinase receptors.

A/Reference number: A35963; PMID:90332672; PMID:2165604

A/Accession: A35963

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-806 <PAS>

A/Cross-references: GB:M35195; NID:g211442; PIDN:AAA48664.1; PID:g211443

C/Genetic: cck2

C/Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prote

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-806/Product: protein-tyrosine kinase cck2 #status predicted <MAT>

F:24-368/Domain: extracellular #status predicted <EXT>

F:131-138/Region: acidic

F:262-335/Domain: immunoglobulin homology <IMM>

F:369-389/Domain: transmembrane #status predicted <TM>

F:390-806/Domain: intracellular #status predicted <INT>

F:464-749/Domain: protein kinase homology <KIN>

F:472-480/Region: protein kinase ATP-binding motif

F:51-107,170-222,269-333/Disulfide bonds: #status predicted

F:96,219,256,288,309,322/Binding site: carboxylate (Asn) (covalent) #status predicted

F:502,519,611/Active site: Lys, Glu, Asp #status predicted

F:616,629/Binding site: magnesium (Asn, Asp) #status predicted

F:642/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 7.6%; Score 100.5; DB 2; Length 806;  
Best Local Similarity 20.8%; Pred. No. 12;  
Matches 59; Conservative 32; Mismatches 85; Indels 107; Gaps 12;

QY 10 LLLVLDLALPANTG-----NKVVLGKGDVETLCTASQKSI 49  
11 LCLAAVAGLPARRRGAERSGGQAAEYRSETAFLELVFG-SDTIELSCN-TQSSSV 68

QY 50 QFHW-----XNSNQIKILNQSGFLTKG 72  
69 SVFWFKDGIAPSNRTHIGQKLKILNVSYDSGLYSCKPRHSN-VLGNTVAVTDS 126

QY 73 PSKLNDRADSRSLMDQGNFPLIKNKLIKEDSDTYICEVEDQKEEVQLVFGILTANSDTH 132  
127 PSSGDEDEDDEDS-----EDTGVFWPTPRDKMEKTLAV--PAAN----- 164

QY 133 LQGSGLTLTLESPPSSPSVQCRS---RSPRKNIQSGKTLVSQEL-----OPS 179  
165 -----TVFRCPAGGNFTPTIYVKNKGERGHRIGIKLRHQGSLWESVPSDR 217

QY 180 GTWCTVQLONOKVE--FKIDIVPRASALPAPTGSALPDPOF 220  
218 GNVTCVENKYNIGNIRRTQDLVLSRPHRIQAG--LPANQT 258

RESULT 93

B42632  
cell adhesion molecule apCAM (clone d15) - California sea hare

C/Species: Aplysia californica (California sea hare)

C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-2000

C/Accession: B42632

R:Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel, E.R.  
Science 256, 638-644, 1992

A/Title: Modulation of an NCAM-related adhesion molecule with long-term synaptic plastic

A/Reference number: A42632; PMID:92263095; PMID:1585176

A/Accession: B42632

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-812 <MAY>

A/Experimental source: CNS

A/Note: sequence extracted from NCBI backbone (NCBIP:101346)

C/Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; imm

Query Match 7.6%; Score 100.5; DB 2; Length 812;  
Best Local Similarity 21.4%; Pred. No. 12;  
Matches 50; Conservative 34; Mismatches 79; Indels 71; Gaps 11;

QY 29 VLGKGDVETLCTASQKSIQIFHKNSNQIKILNQSGFLTKGSKLNDPDRSRSLMD 88  
136 ILGEGE--EVEGVSGKPAFTVTKFENNTKIEAGE--KYTALN----- 176

QY 89 QGNFPLIKNKLIKEDSDTYICEV-----EDQKEEVQLVFGI-----TANSDTHLQ 135  
177 ----KLIKDLSEDTKIKYLCIDIVDTGETKDFIDFTVVKPTALPTIHPDNPV- 231

QY 136 GQSLTLTLESPPSSPSVQCRS---PRKNIQSGKTLVSQELQDSGTWCTV----- 186  
232 GDEVKITCQATGVPPTVQFKGDMVWVTDENVNG-VLTINPLKTTDQATYTCIATNKG 290



Db 6034 NNNCRJPSRDTGKVKVHLEAFISKQKLEIEGKAFVCSISKSEFP-VQMKRD-D 6091

QY 161 KNIQCG-----KTLSSVQLEIDSGTMTCTV 186  
 Db 6092 KTLSSGDKYDVADGKKRVLVVXDATLQDMGTVMV 6128

## RESULT 96

C30127  
 Transmembrane carcinoembryonic antigen 3 precursor - human  
 N:Alternate names: CD66 splice form BQPC  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 23-Jul-1999  
 R:Accession: C30127; 152597  
 R:Barrett, T.R.; Kreuschmer, A.; Austen, D.A.; Goebel, S.J.; Hart, J.T.; Elting, J.J.; J. Cell Biol. 108, 267-276, 1989  
 A:Title: Carcinoembryonic antigens: alternative splicing accounts for the multiple mRNA  
 A:Reference number: A92752; PMID:89139550; PMID:2537311  
 A:Accession: C30127

A:Molecule type: mRNA  
 A:Residues: 1-464 <BAR>  
 A:Cross-references: EMBL:X16356; EMBL:X14784  
 R:Walt, S.M.; Fawcett, J.; Murdoch, S.J.; Teixeira, A.M.; Gschmeissner, S.E.; Hajibagheri  
 Blood 84, 200-210, 1994  
 A:Title: CD66 identifies the biliary glycoprotein (BGP) adhesion molecule: cloning, expr  
 A:Reference number: 152597; PMID:94289702; PMID:8018919

A:Accession: 152597  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-464 <RES>  
 A:Cross-references: GB:571326; NID:9550030; PIDN:AA31183.1; PID:9550031  
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi  
 C:Keywords: alternative splicing; glycoprotein; surface antigen; transmembrane protein  
 F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-464/Product: transmembrane carcinoembryonic antigen 3 #status predicted <MAT>  
 F:160-217/Domain: immunoglobulin homology <IMM1>  
 F:252-301/Domain: immunoglobulin homology <IMM2>  
 F:311-398/Domain: immunoglobulin homology <IMM3>  
 F:44-455/Domain: transmembrane #status predicted <TM>  
 F:104,111,115,152,182,197,208,224,232,254,274,288,292,302,309,345,351,363,378,405/Bindin

Query Match 7.6%; Score 100; DB 2; Length 464;  
 Best Local Similarity 18.6%; Pred. No. 6.8;  
 Matches 49; Conservative 35; Mismatches 78; Indels 102; Gaps 10;

QY 32 KKGDVETCTASQKSIQFMKNSNOIKLNGSFLTKPSKLNDRADSRSLMDQGN 91  
 Db 250 RFGANLTLSCVAASNPRAQYSW-----LNG-----TFQGST 281  
 QY 92 FPLIKRLKIEDSTYICEVDO-----KEVQLVFGLTANS 129  
 Db 282 QEFPIPIYVNSSYCHANNSTVGNRTTKTIVTELSPVAKQIKASKTTVIGDK 341  
 QY 130 DTHLLQOQSLLTLESPGSSPSVQCR-----SPRKNIOGKTLVSQLEIDS 179  
 Db 342 D-----SVNLHST---NDTGISIMFKNQSLPSEEMKLSQNTTISINPKEDA 391  
 QY 180 GTWCTVLQNKQKVEKIDIVPRASALPAPPTGALPDP---QTASALPDPAPASALPA 235  
 Db 392 GTWCEVFN-----PISKQSDPIMLVNVALPQENGLS--PG 428  
 QY 236 ALA-----VTSFLGLGLGAVACVL 254  
 Db 429 AIAIGIVGVALVALVALACFL 452

## RESULT 97

A32164  
 Biliary glycoprotein 1 precursor, splice form a - human  
 N:Alternate names: transmembrane carcinoembryonic antigen 1 (TM1-CEA); transmembrane can  
 N:Contents: biliary glycoprotein 1, splice form b; biliary glycoprotein 1, splice form x  
 C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: A32164; A30127; B30127; A48078; S45664; S65939; A30847; G44476  
 R:Hindoa, Y.; Neumaier, M.; Hefta, S.A.; Dzieniek, Z.; Wagener, C.; Shively, L.; Hefta,  
 Proc. Natl. Acad. Sci. U.S.A. 86, 1668, 1989  
 A:Reference number: A32164  
 A:Contents: extratum

A:Accession: A32164  
 A:Molecule type: mRNA  
 A:Residues: 1-526 <HIN>  
 A:Cross-references: GB:U03858; NID:9179439; PIDN:AAA51826.1; PID:9179440  
 R:Hindoa, Y.; Neumaier, M.; Hefta, S.A.; Dzieniek, Z.; Wagener, C.; Shively, L.; Hefta,  
 Proc. Natl. Acad. Sci. U.S.A. 85, 6959-6963, 1988  
 A:Title: Molecular cloning of a cDNA coding biliary glycoprotein 1: Primary structure o  
 A:Reference number: A94206; PMID:88320555; PMID:2457922

A:Contents: annotation  
 A>Note: the sequence shown in this reference has been completely corrected in reference  
 R:Barrett, T.R.; Kreuschmer, A.; Austen, D.A.; Goebel, S.J.; Hart, J.T.; Elting, J.J.; J.  
 J. Cell Biol. 108, 267-276, 1989  
 A:Title: Carcinoembryonic antigens: alternative splicing accounts for the multiple mRNA  
 A:Reference number: A92752; PMID:89139550; PMID:2537311  
 A:Accession: A30127

A:Molecule type: mRNA  
 A:Residues: 1-526 <BAR>  
 A:Cross-references: EMBL:X16354; NID:937197; PIDN:CAA34404.1; PID:937198; EMBL:X14784  
 A:Experimental source: splice form a  
 A:Accession: B30127

A:Molecule type: mRNA  
 A:Residues: 1-319; 'D', 417-526 <BAR2>  
 A:Cross-references: EMBL:X14831; NID:937199; PIDN:CAA32940.1; PID:937200; EMBL:X14784  
 A:Experimental source: splice form b  
 R:Barrett, T.R.; Drake, L.; Pickle II, W.  
 Mol. Cell. Biol. 13, 1273-1282, 1993  
 A:Title: Human biliary glycoprotein gene: characterization of a family of novel alterna  
 A:Reference number: A48078; PMID:93140765; PMID:8423792  
 A:Accession: A48078

A:Molecule type: mRNA  
 A:Residues: 124-141; 'H', 417-526 <BAR3>  
 A:Cross-references: GB:M76742; NID:9179480; PIDN:AAA57142.1; PID:9179481  
 A:Experimental source: splice form x  
 A>Note: sequence extracted from NCBI backbone (NCBIN:123602, NCBI:123606)  
 A>Note: neither the complete nucleic acid sequence nor the complete translation are sho  
 R:Huck, W.; Neddelic, P.; Turbide, C.; Stanners, C.P.; Barnett, T.R.; Beauchemin, N.  
 Eur. J. Biochem. 223, 529-541, 1994  
 A:Title: Transcriptional control of the human biliary glycoprotein gene, a CEA gene fam

A:Reference number: S45664; PMID:9433343; PMID:8055523  
 A:Accession: S45664  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-21 <HAU>  
 A:Cross-references: EMBL:X67277; NID:929447; PIDN:CAA47694.1; PID:9606777  
 R:Neddelic, P.; Turbide, C.; Beauchemin, N.  
 Eur. J. Biochem. 231, 104-114, 1995  
 A:Title: Characterization and transcriptional activity of the mouse biliary glycoprotei  
 A:Reference number: S65939; PMID:95354678; PMID:7628460  
 A:Accession: S65939  
 A:Status: preliminary; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-21 <NED>  
 A:Cross-references: EMBL:X67277; NID:929447; PIDN:CAA47694.1; PID:9606777  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992  
 A>Note: only a part of the coding sequence is given  
 R:Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstrom, S.  
 Genomics 14, 384-390, 1992  
 A:Title: Identification of three new genes and estimation of the size of the carcinoemb  
 A:Reference number: A44476; PMID:93052339; PMID:1427854  
 A:Contents: annotation; alignment of related sequences  
 C:Genetics:

A:Gene: GDB:BGP  
 A:Cross-references: GDB:127992; OMIM:109770  
 A:Map position: 19q13.2-19q13.2  
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi  
 C:Keywords: alternative splicing; glycoprotein; surface antigen; transmembrane protein  
 F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 3.5436 Seconds  
(without alignments)  
3791.557 Million cell updates/sec

Title: SEQ6  
Perfect score: 1317  
Sequence: 1 MNRGVPRHLLVLQLALLP.....VISPLGLGVACVLARTR 256

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	77.7	458	1	CD4_HUMAN
2	999	75.9	458	1	CD4_PANTR
3	912	69.2	458	1	CD4_MACR1
4	910	69.1	458	1	CD4_MACR1
5	904	68.6	458	1	CD4_MACR1
6	903	68.6	458	1	CD4_MACR1
7	885	67.2	458	1	CD4_MACR1
8	784	59.5	397	1	CD4_CER10
9	783	59.5	397	1	CD4_CER10
10	737.5	56.0	457	1	CD4_SAISC
11	598	45.4	463	1	CD4_CANPA
12	596.5	45.3	459	1	CD4_RABIT
13	490.5	37.2	457	1	CD4_RAT
14	475	36.1	457	1	CD4_MOUSE
15	332	25.2	240	1	CD7_HUMAN
16	133	10.3	1259	1	CAMEL_RAT
17	131	9.9	1260	1	CAMEL_MOUSE
18	129	9.8	338	1	LAMP_HUMAN
19	127	9.6	837	1	NCM2_MOUSE
20	126	9.6	338	1	LAMP_RAT
21	124	9.4	739	1	VCAI_RAT
22	122	9.3	348	1	KILO_RAT
23	120	9.1	349	1	LACH_SCHAM
24	120	9.1	1277	1	CAMEL_FUGRU
25	118	9.0	1010	1	CONT_CHICK
26	118	9.0	1197	1	CAMEL_BRARE
27	117.5	8.9	345	1	OPCM_BOVIN
28	117	8.9	338	1	LAMP_CHICK
29	117	8.9	739	1	VCAI_HUMAN
30	116.5	8.8	345	1	OPCM_HUMAN
31	116.5	8.8	359	1	LACH_DROME
32	112.5	8.5	773	1	PIGR_RABIT
33	111	8.4	702	1	CEAS_HUMAN

34	110.5	8.4	2029	1	LAR_DROME
35	110	8.3	739	1	VCAI_MOUSE
36	109.5	8.3	333	1	AMAL_DROME
37	109	8.3	1257	1	CAMEL_HUMAN
38	109	8.3	1367	1	VGR2_MOUSE
39	109	8.3	4391	1	PGEM_HUMAN
40	108	8.2	257	1	FCEA_HUMAN
41	108	8.2	319	1	A33_HUMAN
42	108	8.2	837	1	NCM2_HUMAN
43	107	8.1	761	1	NCA2_HUMAN
44	107	8.1	848	1	NCAI_HUMAN
45	107	8.1	1447	1	DCC_MOUSE
46	107	8.1	3707	1	PGEM_MOUSE
47	106.5	8.1	345	1	OPCM_RAT
48	105	8.0	2012	1	DSCA_HUMAN
49	105	8.0	344	1	NTRI_HUMAN
50	105	8.0	344	1	NTRI_MOUSE
51	104	7.9	344	1	CONT_HUMAN
52	104	7.9	1018	1	CONT_HUMAN
53	104	7.9	1020	1	CONT_MOUSE
54	104	7.9	1021	1	CONT_RAT
55	103.5	7.9	108	1	KV55_MOUSE
56	103.5	7.9	129	1	KV1W_HUMAN
57	103.5	7.9	210	1	CD7_MOUSE
58	103.5	7.9	1302	1	NRG_DROME
59	103	7.8	519	1	ECTO_RAT
60	103	7.8	521	1	LAC3_MOUSE
61	102.5	7.8	862	1	CD22_MOUSE
62	102.5	7.8	1348	1	VGR2_COTJA
63	102	7.7	521	1	CEAI_MOUSE
64	102	7.7	1343	1	VGR2_RAT
65	101.5	7.7	509	1	SHS1_RAT
66	101.5	7.7	725	1	NCA2_MOUSE
67	101.5	7.7	858	1	NCAI_RAT
68	101.5	7.7	1115	1	NCAI_MOUSE
69	101	7.7	135	1	YOR6_ADEG1
70	101	7.7	1036	1	AXO1_CHICK
71	101	7.7	3375	1	UN52_CAEEL
72	100.5	7.6	108	1	KV50_MOUSE
73	100.5	7.6	337	1	OPCM_CHICK
74	100.5	7.6	806	1	CEK2_CHICK
75	100	7.5	526	1	CEAI_HUMAN
76	98.5	7.5	349	1	CEAB_HUMAN
77	98.5	7.5	686	1	SILB_HUMAN
78	98	7.4	209	1	CD8B_SAISC
79	98	7.4	764	1	PIGR_HUMAN
80	98	7.4	880	1	TYO3_MOUSE
81	97.5	7.4	108	1	KV5T_MOUSE
82	97.5	7.4	322	1	ICOL_MOUSE
83	97.5	7.4	620	1	SMP_COTJA
84	97.5	7.4	771	1	PIGR_MOUSE
85	97.5	7.4	1333	1	VGR1_MOUSE
86	97.5	7.4	6632	1	UN89_CAEEL
87	97	7.4	822	1	FGRI_MOUSE
88	97	7.4	1091	1	NCAI_CHICK
89	97	7.4	1447	1	DCC_HUMAN
90	96.5	7.3	117	1	KV1D_HUMAN
91	96.5	7.3	769	1	PIGR_RAT
92	96.5	7.3	1694	1	SN_MOUSE
93	96	7.3	628	1	LU_HUMAN
94	95.5	7.3	1363	1	VGR3_MOUSE
95	95.5	7.3	1493	1	NEOI_MOUSE
96	95	7.2	245	1	FCER1_RAT
97	95	7.2	1356	1	VGR2_HUMAN
98	94.5	7.2	210	1	KVIP_HUMAN
99	94.5	7.2	108	1	CD8B_PONPY
100	94.5	7.2	524	1	BUTY_MOUSE
101	94	7.1	108	1	KV1S_HUMAN
102	94	7.1	808	1	FGRI_MOUSE
103	94	7.1	822	1	FGRI_HUMAN
104	93.5	7.1	824	1	VGR1_RAT
105	93.5	7.1	824	1	MLTI_HUMAN
106	93.5	7.1	1906	1	KML5_CHICK

P16621	drosofila
P29533	mus musculus
P15364	drosofila
P32004	homo sapien
P35918	mus musculus
P98160	homo sapien
P12319	homo sapien
O99795	homo sapien
O15394	homo sapien
P13592	homo sapien
P13591	homo sapien
P70211	mus musculus
O05793	mus musculus
P32736	rattus norv
O60469	homo sapien
O29121	homo sapien
O99P10	mus musculus
O62718	rattus norv
O12860	homo sapien
P12960	mus musculus
O63198	rattus norv
P01652	mus musculus
P04431	homo sapien
P50283	mus musculus
P20241	drosofila
P16573	rattus norv
O61790	mus musculus
P35329	mus musculus
P52583	coturnix co
P31809	mus musculus
O08775	rattus norv
P97710	protein-t
P13594	mus musculus
P13596	rattus norv
P13595	mus musculus
P20748	avian adeno
P28685	gallus gall
O06551	caenorhabdi
P01650	mus musculus
O98892	gallus gall
P18460	gallus gall
P13668	homo sapien
O91917	homo sapien
O96196	homo sapien
O29877	saimiri sci
P01833	homo sapien
P55144	mus musculus
P01653	mus musculus
O91h38	mus musculus
O92154	coturnix co
O70570	mus musculus
P35959	mus musculus
O01761	caenorhabdi
P01602	mus musculus
P13590	gallus gall
P43166	homo sapien
P01602	homo sapien
P15058	rattus norv
O62230	mus musculus
P50895	homo sapien
P35917	mus musculus
P97798	mus musculus
P12371	rattus norv
P35968	homo sapien
P01608	homo sapien
P30448	pongy pygma
O62556	mus musculus
P01611	homo sapien
O03142	mus musculus
P11362	homo sapien
P53767	rattus norv
O9ud88	homo sapien
P11799	gallus gall

ID	CD4_HUMAN	STANDARD	PRT	458 AA.
107	93	7.1	302	1 Y016_RICPR
108	93	7.1	393	1 ILIS_CERAE
109	93	7.1	853	1 NCAL_BOVIN
110	93	7.1	1338	1 VGR1_HUMAN
111	92.5	7.0	108	1 KVSR_MOUSE
112	92.5	7.0	213	1 CD8B_MOUSE
113	92.5	7.0	344	1 CEA6_HUMAN
114	92.5	7.0	997	1 SCP1_RAT
115	92.5	7.0	1443	1 NEOL_CHICK
116	92	7.0	697	1 SILA_HUMAN
117	91.5	6.9	315	1 HEMA_VACCT
118	91.5	6.9	468	1 IL6A_HUMAN
119	91.5	6.9	738	1 PECL_HUMAN
120	91	6.9	639	1 NRGI_HUMAN
121	90.5	6.9	108	1 KVSM_MOUSE
122	90.5	6.9	335	1 PSG5_HUMAN
123	90.5	6.9	1567	1 FMN2_MOUSE
124	90.5	6.9	1814	1 TSC2_MOUSE
125	90	6.8	265	1 CEA7_HUMAN

## ALIGNMENTS

RESULT 1

CD4\_HUMAN

ID CD4\_HUMAN STANDARD PRT 458 AA.

AC P01730; 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Len-3).

DE CD4.

GN Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85254948; PubMed=2990730;

RA Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chess L., Axel R.;

RT "the isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family.";

RL Cell 42:93-104(1985).

RN [2]

RP REVISION TO 26.

RX MEDLINE=89028665; PubMed=3263213;

RA Littman D.R., Maddon P.J., Axel R.;

RT "Corrected CD4 sequence.";

RL Cell 55:541-541(1988).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=96303695; PubMed=8723724;

RA Asanari-Lari M.A., Muzny D.M., Lu J., Lu F., Lally C.E., Spanos S., Malley T., Gibbs R.A.;

RT "A gene-rich cluster between the CD4 and triosephosphate isomerase genes at human chromosome 12p13.";

RL Genome Res. 6:314-326(1996).

RN [4]

RP SEQUENCE FROM N.A. AND VARIANT TRP-265.

RX MEDLINE=91216786; PubMed=1708753;

RA Hodge T.W., Sasso D.R., McDougal J.S.;

RT "Humans with OKT4-epitope deficiency have a single nucleotide base change in the CD4 gene, resulting in substitution of TRP240 for ARG240.";

RL Hum. Immunol. 30:99-104(1991).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altmann S.F., Zeeberg B., Buecaw K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Martusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smailus D.E., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A., Scherch A., Schein J.E., Jones S.J.W., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [6]

RP SEQUENCE OF 28-424 FROM N.A.

RX TISSUE=Blood;

RC MEDLINE=93049640; PubMed=1425921;

RA Fomsgaard A., Hirsch V.M., Johnson P.R.;

RT "Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus.";

RL Eur. J. Immunol. 22:2973-2981(1992).

RN [7]

RP SEQUENCE OF 26-394.

RX MEDLINE=90078232; PubMed=2592374;

RA Carr S.A., Henling M.E., Folea-Wasserman G., Sweet R.W., Anumula K., Barr J.R., Huddleston M.J., Taylor P.;

RT "Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry.";

RL J. Biol. Chem. 264:21286-21295(1989).

RN [8]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.

RX MEDLINE=91061861; PubMed=1701030;

RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L., Tarr G.E., Husein Y., Reinherz E.L., Harrison S.C.;

RT "Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains.";

RL Nature 348:411-418(1990).

RN [9]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.

RX MEDLINE=91061882; PubMed=2247146;

RA Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Arthos J., Rosenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.;

RT "Crystal structure of an HIV-binding recombinant fragment of human CD4.";

RL Nature 348:419-426(1990).

RN [10]

RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.

RX MEDLINE=97311402; PubMed=9168119;

RA Wu H., Kwong P.D., Hendrickson W.A.;

RT "Dimeric association and segmental variability in the structure of human CD4.";

RL Nature 387:527-530(1997).

RN [11]

RP PALMITOYLATION.

RX MEDLINE=92317088; PubMed=1618861;

RA Chise B., Rose J.K.;

RT "Identification of palmitoylation sites on CD4, the human immunodeficiency virus receptor.";

RL J. Biol. Chem. 267:13593-13597(1992).

CC -I- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.

CC -I- SUBUNIT: Associates with p56-lck.

CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

CC -I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domain.

CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.



```

CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD4 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd4.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12807; AAA3572.1; .
DR EMBL; U47924; AAB51309.1; .
DR EMBL; M35160; AAA16069.1; .
DR EMBL; BC025782; AAB25782.1; .
DR PIR; A90872; RWHUT4.
DR PDB; 1CDH; 30-APR-94.
DR PDB; 1CDI; 30-APR-94.
DR PDB; 3CD4; 31-OCT-93.
DR PDB; 1CDU; 01-APR-97.
DR PDB; 1CDU; 01-APR-97.
DR PDB; 1CDY; 01-APR-97.
DR PDB; 1WBR; 12-MAR-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1G9M; 27-DEC-00.
DR PDB; 1G9N; 27-DEC-00.
DR PDB; 1G11; 19-SEP-01.
DR PDB; 1JL4; 19-SEP-01.
DR GlycoSiteDB; P01730; .
DR GeneW; HGNC:1678; CD4.
DR MIM; 186940; .
DR GO; GO:0042101; C:T-cell receptor complex; NAS.
DR GO; GO:0015026; F:coreceptor activity; NAS.
DR GO; GO:0015029; F:internalization receptor activity; TAS.
DR GO; GO:0042289; F:MHC class II protein binding; NAS.
DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR GO; GO:0009405; P:pathogenesis; TAS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; NAS.
DR GO; GO:0030217; P:T-cell differentiation; NAS.
DR GO; GO:0045058; P:T-cell selection; NAS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; NAS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PRO0692; CD4TCANTIGEN.
DR SMART; SMO0406; IGV_1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
KW Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSLEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT VARIANT 265 265

```

```

FT STRAND 27 32 /FtId=VAR_003906.
FT TURN 33 34
FT STRAND 37 39
FT TURN 44 45
FT STRAND 51 55
FT TURN 56 57

Query Match 77.7%; Score 1023; DB 1; Length 458;
Best Local Similarity 99.5%; Pred. No. 1,1e-70;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRLHLLVQLALPAATQGNKVVGGKDYELCTASQSKSIQFMHNSNQIK 60
Dd 1 MNRGVPFRLHLLVQLALPAATQGNKVVGGKDYELCTASQSKSIQFMHNSNQIK 60
QY 61 ILGNQGSFLTKGSPSKINDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
Dd 61 ILGNQGSFLTKGSPSKINDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLQGSLLTLTSPGSSPSVQCRRPRKNIQGGKTLISVSOLELDSG 180
Dd 121 LVFGLTANSDTHLLQGSLLTLTSPGSSPSVQCRRPRKNIQGGKTLISVSOLELDSG 180
QY 181 TWICTVQONOKVFEKIDIV 200
Dd 181 TWICTVQONOKVFEKIDIV 200

RESULT 2
CD4_PANTR
ID CD4_PANTR STANDARD; PRT; 458 AA.
AC P16004;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Lew-3).
GN CD4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9018264; PubMed=2107024;
RA Camerini D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE OF 26-424 FROM N.A.
RX TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

C      EMBL; M31135; AAA35407.1; -.
DR      EMBL; X73723; CAA51749.1; -.
DR      PIR; B32722; RWCZT4.
DR      HSSP; P01730; 1WQ.
DR      GO; GO:0042101; C:T-cell receptor complex; ISS.
DR      GO; GO:0015026; F:coreceptor activity; ISS.
DR      GO; GO:0042289; F:MHC class II protein binding; ISS.
DR      GO; GO:0006955; P:immune response; ISS.
DR      GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR      GO; GO:0030217; P:T-cell differentiation; ISS.
DR      GO; GO:0045058; P:T-cell selection; ISS.
DR      GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR      InterPro; IPR000973; CD4 TCAG.
DR      InterPro; IPR007110; Ig_V.
DR      Pfam; PF00047; Ig_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PSS0835; IG_LIKE; 1.
KW      Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KM      Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT      SIGNAL 1 25
FT      CHAIN 26 458
FT      DOMAIN 26 396
FT      TRANSMEM 397 418
FT      DOMAIN 419 458
FT      DOMAIN 26 125
FT      DOMAIN 126 203
FT      DOMAIN 204 317
FT      DOMAIN 318 374
FT      CARBOHYD 296 296
FT      CARBOHYD 325 325
FT      DISULFID 41 109
FT      DISULFID 155 184
FT      DISULFID 328 370
FT      LIPID 419 419
FT      LIPID 422 422
FT      CONFLICT 42 42
FT      CONFLICT 62 62
FT      CONFLICT 191 191
SQ      SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;

Query Match 75.9%; Score 999; DB 1; Length 458;
Best Local Similarity 97.5%; Pred. No. 7.6e-69;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
DB 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
QY 61 ILGNQGSFLTKGSPKLNDRADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVREGLTANSPHTLQSGSLTLTLESPGSSPVQCRSPKGNIOGKTLISVSOLELDQSG 180
DB 121 LVREGLTANSPHTLQSGSLTLTLESPGSSPVQCRSPKGNIOGKTLISVSOLELDQSG 180
QY 181 TWICTVLOKQKVEFKIDIV 200
DB 181 TWICTVLOKQKVEFKIDIV 200
QY 181 TWICTVLOKQKVEFKIDIV 200
DB 181 TWICTVLOKQKVEFKIDIV 200

RESULT 3
CD4_MACFA STANDARD; PRT; 458 AA.
AC P79185;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3).

```

```

GN      CD4.
OS      Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC      Cercopithecinae; Macaca.
OX      NCBI_TaxId=9541;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Thymocytes;
RA      Tatemui M., Yabe M., Yamada Y.K.;
RL      Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC      receptor interaction. May regulate T-cell activation.
CC      -1- SUBUNIT: Associates with p56-lck (By similarity).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC      -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.ch).
C      EMBL; D63349; BAA09673.1; -.
DR      HSSP; P01730; 1WBR.
DR      GO; GO:0042101; C:T-cell receptor complex; ISS.
DR      GO; GO:0015026; F:coreceptor activity; ISS.
DR      GO; GO:0042289; F:MHC class II protein binding; ISS.
DR      GO; GO:0006955; P:immune response; ISS.
DR      GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR      GO; GO:0030217; P:T-cell differentiation; ISS.
DR      GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR      InterPro; IPR000973; CD4 TCAG.
DR      InterPro; IPR007110; Ig_Like.
DR      Pfam; PF00047; Ig_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PSS0835; IG_LIKE; 1.
KW      Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KM      Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT      SIGNAL 1 25
FT      CHAIN 26 458
FT      DOMAIN 26 396
FT      TRANSMEM 397 418
FT      DOMAIN 419 458
FT      DOMAIN 26 125
FT      DOMAIN 126 203
FT      DOMAIN 204 317
FT      DOMAIN 318 374
FT      CARBOHYD 296 296
FT      CARBOHYD 325 325
FT      DISULFID 41 109
FT      DISULFID 155 184
FT      DISULFID 328 370
FT      LIPID 419 419
FT      LIPID 422 422
SQ      SEQUENCE 458 AA; 50872 MW; 9105479FB5656FF7 CRC64;

Query Match 69.2%; Score 912; DB 1; Length 458;
Best Local Similarity 88.0%; Pred. No. 3.2e-62;
Matches 176; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
DB 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
QY 61 ILGNQGSFLTKGSPKLNDRADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120

```

```

DB 61 ILGIQGSFLTKGPKSLSDRADSRKSLMDQGCFSMTIKNLKIEDSDTYICEVANKKEEVEL 120
OY 121 LVFGILTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLEGSLTLTLESPPSSPSVKCRSPGKNIQCGRTISVPLERDQSG 180
OY 181 TWCTVLONOKKVEFKIDIV 200
DB 181 TWCTVSDQKTVFKIDIV 200

RESULT 4
CD4_MACFU STANDARD; PRT: 458 AA.
AC P79184.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Macaca fasciata fasciata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; D63348; BAA09672.1; -.
DB HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; IG_V.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25 T-C-CELL SURFACE GLYCOPROTEIN CD4.
FT CHAIN 26 458 BY SIMILARITY.
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 418 POTENTIAL.
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.

```

```

FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 458 AA; 50828 MW; 7653E7EF0818535 CRC64;

Query Match 69.1%; Score 910; DB 1; Length 458;
Best Local Similarity 87.5%; Pred. No. 4.6e-62;
Matches 175; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

OY 1 MNRGVPFRHLILVQLALPPATQGNKYVLAKKDDYELTCTASQKKSIOFHMKNSQIK 60
DB 1 MNRGIPRHLILVQLALPPAVTQKKYVLAKKDDYELTCTASQKKNTOFHMKNSQIK 60
OY 61 ILGIQGSFLTKGPKSLSDRADSRKSLMDQGCFSMTIKNLKIEDSDTYICEVANKKEEVEL 120
DB 61 ILGIQGSFLTKGPKSLSDRADSRKSLMDQGCFSMTIKNLKIEDSDTYICEVANKKEEVEL 120
OY 121 LVFGILTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLEGSLTLTLESPPSSPSVKCRSPGKNIQCGRTISVPLERDQSG 180
OY 181 TWCTVLONOKKVEFKIDIV 200
DB 181 TWCTVSDQKTVFKIDIV 200

RESULT 5
CD4_MACMU STANDARD; PRT: 458 AA.
ID CD4_MACMU
AC P16003; Q29617;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Camerini D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Hashimoto O., Tatsumi M.;
RT Molecular cloning and expression of macaque CD4s.";
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [4]
RP SEQUENCE OF 107-192 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;

```

```

RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -|- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -|- SUBUNIT: Associates with p56-lck (By similarity).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31134; AAA56838.1; -
DR EMBL; D63347; BAA09670.1; -
DR EMBL; X73326; CAA51751.1; -
DR EMBL; AF057385; AAC25129.1; -
DR HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 418 POTENTIAL.
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 N -> T (IN REF. 1).
FT CONFLICT 42 42 L -> S (IN REF. 3).
FT CONFLICT 62 62 L -> S (IN REF. 2).
FT CONFLICT 67 67 I -> L (IN REF. 2).
FT CONFLICT 169 169 K -> N (IN REF. 3).
FT CONFLICT 191 191 S -> P (IN REF. 2).
FT CONFLICT 248 248 R -> Q (IN REF. 3).
FT CONFLICT 265 265 A -> T (IN REF. 2).
FT CONFLICT 349 349
SQ SEQUENCE 458 AA; 50884 MW; 8BB80339FAFC808 CRC64;

Query Match 68.6%; Score 904; DB 1; Length 458;
Best Local Similarity 87.0%; Pred. No. 1,3e-61;
Matches 174; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 61 IIGNOSFLTKGPKLNDADSRSLMDGNFLLIKNLIKIDSDPTICEVEDQKEVOL 120
DB 61 IIGIOQLFTLKPKSKLSDADSRKSLMDQGCFSMTIKNIKIEDSDPTICEVEKKEVEL 120
QY 121 LVFGLTANSDTHLQGSITLTLTLESPGSSPVOCSPRGKNIQGSKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQGSITLTLTLESPGSSPVOCSPRGKNIQGSRTISVPLERQDSG 180
QY 181 TWCTVLTQNKYVERKIDIV 200
DB 181 TWCTVLTQNKYVERKIDIV 200

RESULT 6
CD4_MACNE STANDARD; PRT; 458 AA.
ID CD4_MACNE
AC Q08340; P79196;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Macaca nemestrina (pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9545;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsu M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RA TISSUE=Blood;
RC MEDLINE=93049640; PubMed=1425921;
RA Pomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -|- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -|- SUBUNIT: Associates with p56-lck (By similarity).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63346; BAA09670.1; -
DR EMBL; X73325; CAA51751.1; -
DR HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.

```



```

FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
FT CONFLICT 46 46 K -> N (IN REF. 2 AND 3; AAB60875).
FT CONFLICT 59 59 I -> T (IN REF. 1).
FT CONFLICT 115 115 K -> E (IN REF. 3; AAB60873).
FT CONFLICT 165 165 G -> V (IN REF. 3; AAB60873 AND 4).
FT CONFLICT 200 200 M -> L (IN REF. 2; AAB60873).
FT CONFLICT 227 227 F -> L (IN REF. 3; AAB60873).
FT CONFLICT 271 271 K -> E (IN REF. 3; AAB60873).
FT CONFLICT 281 281 N -> H (IN REF. 3; AAB60873).
SQ SEQUENCE 458 AA; 51158 MW; FC523D2EDD1F72E7 CRC64;

Query Match 67.2%; Score 885; DB 1; Length 458;
Best Local Similarity 85.5%; Pred. No. 3,7e-60;
Matches 171; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 1 MNRGVPRLHLLVQLALPAATQGNKVVIGKGGDTVELCTASQKKSIOFHKNNOIK 60
DB 1 MNRGVPRLHLLVQLALPAATQGNKVVIGKGGDTVELCTASQKKSIOFHKNNOIK 60
QY 61 ILGNQGSFLTKGSKLNDRAISRSLMDQGNPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGSKLNDRAISRSLMDQGNPLIINKLIEDSDTYICEVEDQKEVOL 120
QY 121 LVRLGINSPTHLIOGSLTLTLSPGSSPSVQCRSPRGKNTGGKTLVSQLELDSG 180
DB 121 LVRLGINSPTHLIOGSLTLTLSPGSSPSVQCRSPRGKNTGGKTLVSQLELDSG 180
QY 181 TWICTVONOKKVEFKIDIV 200
DB 181 TWICTVONOKKVEFKIDIV 200
QY 181 TWICTVONOKKVEFKIDIV 200
DB 181 TWICTVONOKKVEFKIDIV 200

RESULT 8
ID_CD4_CERTO STANDARD; PRT; 397 AA.
AC 008336;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
DB (Fragment).
OS Cercopithecus torquatus alyx (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_Taxid=9531;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RT J. Immunol. 22:2973-2981(1992).
RL BUR. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announc/

```

```

CC or send an email to license@sb-sib.ch).
CC EMBL: X73328; CAAS1754.1; -.
CC EMBL: X73327; CAAS1753.1; -.
CC HSSP; P01730; 1MIQ.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0005955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR009973; CD4 TCAG.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1 1
FT DOMAIN <1 369 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT DOMAIN 392 >397 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT LIPID 301 343 S-palmitoyl cysteine (By similarity).
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT VARIANT 20 20 MISSING.
FT VARIANT 43 43 T -> I.
FT VARIANT 86 86 N -> D.
FT VARIANT 96 96 F -> L.
FT VARIANT 173 173 V -> M.
FT VARIANT 316 316 R -> K.
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2DB38A7 CRC64;

Query Match 59.5%; Score 784; DB 1; Length 397;
Best Local Similarity 87.3%; Pred. No. 1.5e-52;
Matches 151; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 28 VVLGKGGDTVELCTASQKKSIOFHKNNOIKILGNQGSFLTKGSKLNDRAISRSLW 87
DB 1 VVLGKGGDTVELCTASQKKSIOFHKNNOIKILGNQGSFLTKGSKLNDRAISRSLW 87
QY 88 DQGNFPLIINKLIEDSDTYICEVEDQKEVOLVGLTRANSPTHLIOGSLTLTLSP 147
DB 61 DQGNFPLIINKLIEDSDTYICEVEDQKEVOLVGLTRANSPTHLIOGSLTLTLSP 120

RESULT 9
ID_CD4_ERYPA STANDARD; PRT; 397 AA.
AC 008339;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
DB (Fragment).

```

OS CD4.  
 NS Erythrocybus patas (Red guenon) (Huscar).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecidae; Erythrocybus.  
 OX NCBI\_TaxId=9538;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=93049640; PubMed=1425921;  
 RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
 RT "Cloning and sequences of primate CD4 molecules: diversity of the  
 RT cellular receptor for simian immunodeficiency virus/human  
 RT immunodeficiency virus.";  
 RL Eur. J. Immunol. 22:2973-2981(1992).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.  
 CC -1- SUBUNIT: Associates with p56-lck (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X73324; CAA51750.1; -.  
 DR HSBP; P01730; IMIO.  
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.  
 DR GO; GO:0015026; F:coreceptor activity; ISS.  
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.  
 DR GO; GO:0006955; P:immune response; ISS.  
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.  
 DR GO; GO:0030217; P:T-cell differentiation; ISS.  
 DR GO; GO:0045058; P:T-cell selection; ISS.  
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
 DR InterPro: IPR000973; CD4 TCAG.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003566; IG\_v.  
 DR Pfam; PF00047; Ig; 2.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KW Immune response; Repeat; Lipoprotein; Palmitate.  
 KM  
 FT NON\_TER 1  
 FT DOMAIN 1  
 FT TRANSMEM 370 369  
 FT DOMAIN 391  
 FT DOMAIN 392 >397  
 FT DOMAIN <1 98  
 FT DOMAIN 99  
 FT DOMAIN 177 176  
 FT DOMAIN 291 347  
 FT CARBOHYD 269 269  
 FT CARBOHYD 298 298  
 FT DISULFID 14 82  
 FT DISULFID 128 157  
 FT LIPID 301 343  
 FT LIPID 392 392  
 FT LIPID 395 395  
 FT NON\_TER 397 397  
 SO SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;  
 Query Match 59.5%; Score 783; DB 1; Length 397;  
 Best Local Similarity 86.7%; Pred. No. 1.8e-52;  
 Matches 150; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Db	1	VVLAKGDVTVELTGNASOKTTTQGHMKNNSNMKTLGQGSFLTGTGPKLMDRDRSRSLW	60
Qy	88	DQGNPFLTIKLIKLTEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLQGOSLTLTLESPP	147
Db	61	DQGFMSIIRIKLIKLTEDSEYICEVEDKKEVEYELLVFGLTANSDTHLLQGOSLTLTLESPP	120
Qy	148	GSSSVQCRSPRGNIQGKTLVSQLELDDSGTWTCTVLCNKKVEFKIDIV	200
Db	121	GSSSVQCRSPRGNIQGKTLVSQLELDDSGTWTCTVLCNKKVEFKIDIV	173
RESULT 10			
CD4_SAISC			
ID	CD4_SAISC	STANDARD;	PRT; 457 AA.
AC	Q29037;		
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Les-3).		
GN	CD4.		
OS	Saimiri sciureus (Common squirrel monkey).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Euhetia; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.		
OX	NCBI_TaxID=9521;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RA	Tatsumi M., Hashimoto O.;		
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.		
CC	-1- SUBUNIT: Associates with p56-lck (by similarity).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.		
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; D86588; BAA13131.1; -.		
DR	HSSP; P01730; 1WB.		
DR	GO; GO:0042101; C:T-cell receptor complex; ISS.		
DR	GO; GO:0015026; F:coreceptor activity; ISS.		
DR	GO; GO:0042289; F:MHC class II protein binding; ISS.		
DR	GO; GO:0006955; P:immune response; ISS.		
DR	GO; GO:0045066; P:positive regulation of interleukin-2 biosyn. ; ISS.		
DR	GO; GO:0030217; P:T-cell differentiation; ISS.		
DR	GO; GO:0045058; P:T-cell selection; ISS.		
DR	GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.		
DR	InterPro; IPR000973; CD4_FcAg.		
DR	InterPro; IPR007110; IG_Like.		
DR	InterPro; IPR003596; IG_V.		
DR	Pfam; PF00047; IG_2.		
DR	PRINTS; PR00692; CD4TCANTIGEN.		
DR	SMART; SM00406; IGV; 1.		
DR	PROSITE; PS50835; IG_LIKE; 1.		
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell1;		
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate.		
FT	SIGNAL	1	25
FT	CHAIN	26	457
FT	DOMAIN	26	395
FT	TRANSHEM	396	417
FT	DOMAIN	418	457
FT	DOMAIN	26	125
FT	DOMAIN	126	202
FT	DOMAIN	203	316
FT	DOMAIN	317	373
FT	CARBOHYD	254	254
FT	N-LINKED (GLCNAC. .) (POTENTIAL).		

```

FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 327 369 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
SO SEQUENCE 457 AA; 50871 MW; 57EB6344005A015 CRC64;

Query Match 56.0%; Score 737.5; DB 1; Length 457;
Best Local Similarity 71.5%; Pred. No. 6,2e-49;
Matches 143; Conservative 24; Mismatches 32; Indels 1; Gaps 1;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVLGKGGDYELTCTASQKSIQFHKNSNQIK 60
DB 1 MNGGIPFRHLVLVQLALPAATQGNKVLGKGGDYELTCTASQKSIQFHKNSNQIK 60
QY 61 ILGNQGSFLLTKGPKSKINDRSDRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGVQNFVTRGQSKLTDRISSKSSMDRSGFPLIDAIEDSETTYICEVESKEEVEL 120
QY 121 LVFGLTNSDTHLLQGSLLTLTLESPGSSPSVQCSPRGKNIQGGKTLSSVQLQLDQSG 180
DB 121 QVGLGFANPDTHLLQGSLLTLTLESPGSSPSVECTSPRGKIRGRKTLSSVQLGIPDSG 180
QY 181 TWCTVQGNOKVYEFKIDIV 200
DB 181 TWKCTVQGHLELV-FEINIV 199

RESULT 11
ID CD4_CANFA STANDARD; PRT; 463 AA.
AC P33705;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE OF 13-463 FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=93192324; PubMed=7916632;
RA Milde K.F., Conner G.E., Minz D.H., Alejandro R.;
RT "Primary structure of the canine CD4 antigen.";
RL Biochim. Biophys. Acta 1172:315-318(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
T lymphocytes.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L06130; AAB02295.1; -.
DR EMBL; X68565; -, NOT_ANNOTATED_CDS.
DR HSSP; P01730; 1WBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_3.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 24
FT CHAIN 25 463
FT DOMAIN 25 401
FT TRANSMEM 402 423
FT DOMAIN 424 463
FT DOMAIN 26 124
FT DOMAIN 125 211
FT DOMAIN 212 321
FT DOMAIN 322 378
FT DISULFID 41 109
FT DISULFID 332 374
FT LIPID 424 424
FT LIPID 427 427
FT CARBOHYD 123 123
FT CARBOHYD 168 168
FT CARBOHYD 176 176
FT CARBOHYD 324 324
FT CARBOHYD 329 329
FT CARBOHYD 389 389
SO SEQUENCE 463 AA; 51639 MW; 95805170CB44A833 CRC64;

Query Match 45.4%; Score 598; DB 1; Length 463;
Best Local Similarity 55.5%; Pred. No. 2,7e-38;
Matches 116; Conservative 38; Mismatches 45; Indels 10; Gaps 2;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVLGKGGDYELTCTASQKSIQFHKNSNQIK 60
DB 1 MNGGAFPRHLMLQLVLMIPATVPRAEVYLGKAGDAVELPCQTSQKNIHNRDSSMWQ 60
QY 61 ILGNQGSFLLTKGPKSKINDRSDRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLLTKGPKSKINDRSDRSRLMDQGSFPLVIXDLEVDGSIYFCDT-DKQEEVEL 119
QY 121 LVFGLTNSDTHLLQGSLLTLTLESPGSSPSVQCSPRGKNIQGGKTLSSVQLQLDQSG 180
DB 121 LVFGLTNSDTHLLQGSLLTLTLESPGSSPSVQCSPRGKNIQGGKTLSSVQLQLDQSG 180
QY 172 SOLELDGSGTWCTVQGNOKVYEFKIDIV 200
DB 172 SWPELDGSGTWCTVQGNOKVYEFKIDIV 208

RESULT 12
ID CD4_RABIT STANDARD; PRT; 459 AA.
AC P46330;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen

```



```

DE T4/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92390370; PubMed=1518821;
RA Hague B.F., Sawadkrosol S., Brown T.J., Lee K., Recker D.P.,
RA Kindt T.J.;
RT "CD4 and its role in infection of rabbit cell lines by human
RT immunodeficiency virus type 1."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92840; AAA31198.1; -.
DR PIR; A46254; A46254.
DR HSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007159; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig-Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG_V.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 459
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 397 419 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 420 459 POTENTIAL.
FT DOMAIN 26 129 IG-LIKE V-TYPE.
FT DOMAIN 130 208 IG-LIKE C2-TYPE 1.
FT DOMAIN 209 318 IG-LIKE C2-TYPE 2.
FT DOMAIN 319 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 113 POTENTIAL.
FT DISULFID 329 370 POTENTIAL.
FT LIPID 420 420 S-palmitoyl cysteine (By similarity).
FT LIPID 423 423 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 459 AA; 50886 MW; B32311CBDA0013D CRC64;
Query Match 45.3%; Score 596.5; DB 1; Length 459;
Best Local Similarity 57.9%; Pred. No. 3,4e-38;
Matches 124; Conservativity 35; Mismatches 46; Indels 9; Gaps 3;
OY 1 MNRGVPFRHLLVLTQALLPAATGKNNKVLAKGKDPVLELTCTASQKKSIOFHWNKSNQIK 60
DB 1 MNRRIYFCCLLVLPALLPAATGKTVRGKAGAIIVELPCOSSQKSNVFNWGHANQVK 60

```

```

OY 61 ILNGQG-----SFLTGSPSKINDRADSRRLMDQGNFLLIKNLKIEDSPYICEVEDQKE 116
DB 61 ILNGQSSSSSFWLKGNSPLSNRVESKKNMMQDSFFLVIKDLRMDSGTYICEVGVK 120
OY 117 EVQLVFGLTSPDTHLLOQSSTLTLTSPGSSPSVQCSPPKKNIGGKTTLSVSOLEL 176
DB 121 EVELLVFRLTANPTRLHQQSLTLTLTGCPSPSVQVMSPEKKIETGPTCSMPKRL 180
OY 177 ODGCTWCTV-LDNQKVEFEKIDV----PRASA 205
DB 181 QDSGTWSCHLSFDQNTKLELDIKITLVGFPRASA 214
RESULT 13
ID CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3) (M3/25 antigen).
GN CD4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=8717535; PubMed=1104900;
RA Clark S.J., Jefferies W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (M3/25) antigen:
RT evidence for derivation from a structure with four
RT immunoglobulin-related domains."
RL Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RA Brady R.L., Dodson E.D., Dodson G.G., Lange G., Davis S.J.,
RA Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
RT NH2-terminal domains."
RL Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15768; AAA40901.1; -.
DR PIR; A27449; A27449.
DR PDB; 1CID; 15-JUL-93.
DR GlycoStated; P05540; -.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003599; Ig.

```

DR Pfam: PF00047; i9; 2.  
 DR PRINTS; PRO0692; CD4TCANTIGEN.  
 DR SMART; SM00409; IG; 2.  
 DR PROSITE; PSS00835; IG-LIKE; 1.  
 KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; 1.  
 KM Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.  
 FT SIGNAL 1 27  
 FT CHAIN 28 457  
 FT DOMAIN 28 394  
 FT TRANSMEM 395 417  
 FT DOMAIN 418 457  
 FT DOMAIN 28 127  
 FT DOMAIN 128 206  
 FT DOMAIN 207 316  
 FT DOMAIN 317 374  
 FT CARBOHYD 186 186  
 FT CARBOHYD 297 297  
 FT CARBOHYD 392 392  
 FT DISULFID 43 111  
 FT DISULFID 158 187  
 FT DISULFID 328 370  
 FT LIPID 418 418  
 FT LIPID 421 421  
 FT STRAND 213 217  
 FT TURN 218 219  
 FT STRAND 222 225  
 FT STRAND 235 243  
 FT STRAND 252 258  
 FT TURN 259 260  
 FT STRAND 261 265  
 FT STRAND 274 275  
 FT STRAND 278 278  
 FT TURN 279 280  
 FT STRAND 282 285  
 FT HELIX 290 292  
 FT STRAND 294 301  
 FT STRAND 306 319  
 FT STRAND 325 331  
 FT STRAND 338 344  
 FT TURN 345 346  
 FT STRAND 349 353  
 FT STRAND 357 361  
 FT STRAND 367 374  
 FT TURN 375 376  
 FT STRAND 377 385  
 SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;  
 Query Match 37.2%; Score 490.5; DB 1; Length 457;  
 Best Local Similarity 48.8%; Pred. No. 4e-30;  
 Matches 102; Conservative 35; Mismatches 67; Indels 5; Gaps 3;

DT 01-JAN-1988 (Rel. 06, last sequence update)  
 DT 15-MAR-2004 (Rel. 43, last annotation update)  
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen  
 DE T4/Leu-3) (T-cell differentiation antigen L3T4).  
 GN CD4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX MEDLINE=87018845; PubMed=3094146;  
 RX Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;  
 RT "Isolation and sequence of L3T4 complementary DNA clones: expression  
 RT in T cells and brain.";  
 RL Science 234:610-614(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87115821; PubMed=3027575;  
 RX Littman D.R., Gettner S.N.;  
 RT "Unusual intron in the immunoglobulin domain of the newly isolated  
 RT murine CD4 (L3T4) gene.";  
 RL Nature 325:453-455(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88152875; PubMed=3326818;  
 RX Parnes J.R., Hunkapiller T.;  
 RT "L3T4 and the immunoglobulin gene superfamily: new relationships  
 RT between the immune system and the nervous system.";  
 RL Immunol. Rev. 100:109-127(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=88041159; PubMed=2823269;  
 RX Gorman S.D., Tourville B., Parnes J.R.;  
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript  
 RT in brain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98112780; PubMed=9445485;  
 RX Ansari-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,  
 RA Lu J., Gorrell J.H., Chinnault A.C., Belmont J.W., Miller W.,  
 RA Gibbs R.A.;  
 RT "Comparative sequence analysis of a gene-rich cluster at human  
 RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";  
 RL Genome Res. 8:29-40(1998).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22368257; PubMed=12477932;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshlyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,  
 RA Bask S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]

RX		SEQUENCE OF 27-43.
RP	MEDLINE=8616694; PubMed=3082751;	
RA	Claesson B.J., Teagardot U., Kirsbaum L., Maddox J., McKay C.R.,	
RA	Bratton M., McKenzie I.F.C., Walker I.D.;	
RT	"The LT $\alpha$ antigen in mouse and the sheep equivalent are immunoglobulin-like."	
RT	Immunogenetics 23:129-132(1986).	(8)
RV	[8]	
RP	DISULFIDE BONDS:	
RX	MEDLINE=86233454; PubMed=3086886;	
RA	Claesson B.J., Teagardot U., McKenzie I.F.C., Walker I.D.;	
RT	"Partial primary structure of the T $\delta$ antigens of mouse and sheep: assignment of intrachain disulfide bonds;"	
RL	Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503(1986).	
CC	-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.	
CC	-1- SUBUNIT: Associates with p56-lck (By similarity)."	
CC	-1- INTRACELLULAR LOCATION: Type I membrane protein.	
CC	-1- ALTERNATIVE PRODUCTS:	
CC	Event-Alternative splicing; Named isoforms=2;	
CC	Name=1;	
CC	IsoId=P06332-1; Sequence=Displayed;	
CC	Name=2; Synonyms=Brain-specific;	
CC	IsoId=P06332-2; Sequence-VSP_002489;	
CC	-1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.	
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement ( <a href="http://www.isb.ch/announce/">See http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ) .	
CC	----	
DR	EMBL; M36850; AAA39401.1; --	
DR	EMBL; M13816; AAA37267.1; --	
DR	EMBL; X04836; CAA28539.1; --	
DR	EMBL; M36851; AAA39402.1; --	
DR	EMBL; M17080; AAA37403.1; --	
DR	EMBL; M17078; AAA37403.1; JOINED.	
DR	EMBL; M17079; AAA37403.1; JOINED.	
DR	EMBL; AC002397; AAC36010.1; --	
DR	EMBL; BC039137; AAH39137.1; --	
DR	PIR; A02110; RWMST4.	
DR	HSSP; P01730; IMBR.	
DR	GMD; MG1.8835; CD4.	
DR	GO; GO:0042101; C:T-cell receptor complex; ISS.	
DR	GO; GO:0015026; F:coreceptor activity; ISS.	
DR	GO; GO:0042289; F:MHC class II protein binding; ISS.	
DR	GO; GO:0006955; P:immune response; ISS.	
DR	GO; GO:0045066; P:positive regulation of interleukin-2 biosyn.. ; ISS.	
DR	GO; GO:0030217; P:T-cell selection; ISS.	
DR	GO; GO:0045085; P:T-cell selection; ISS.	
DR	InterPro; IPRO00973; CdM TCAG.	
DR	InterPro; IPRO07110; IG-I-like.	
DR	InterPro; IPRO03596; Ig_V.	
DR	Pfam; PF00047; Ig_2.	
DR	PRINTS; PRO0632; CDATCANTIGEN.	
DR	SMART; SM00406; IGV; 1.	
KW	PROSITE; PS50835; IG_LIKE; 1.	
KV	immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;	
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate;	
KV	Alternative splicing	
FT	SIGNAL	1 .. 26
FT	CHAIN	27 .. 457
FT	DOMAIN	27 .. 394 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	395 .. 417 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	418 .. 457 POTENTIAL.
FT	DOMAIN	27 .. 128 IG-LIKE V-TYPE.
FT	DOMAIN	129 .. 207 IG-LIKE C2-TYPE 1.
FT	DOMAIN	208 .. 317 IG-LIKE C2-TYPE 2.

Query Match	Best Local Similarity	Score 475; DB 1; Length 457; Pred. No. 6,1e-29;	Mismatches 55; Indels 6; Gaps 5;
Matches 104; Conservative 32;			
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
FT DOMAIN 318 374		IG-LIKE C2-TYPE 3	
FT CARBOHYD 187 187		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 298 298		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 323 323		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 392 392		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT DISULFID 42 112			
FT DISULFID 159 188			
FT DISULFID 328 370			
FT LIPID 418 418		S-palmitoyl cysteine (By similarity).	
FT LIPID 421 421		S-palmitoyl cysteine (By similarity).	
FT VARSPDIC 1 240		Missing (in isoform 2).	
FT VARSPDIC 1 240		/FTID=VSP 002489.	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
Query Match	36.1%;	Score 475; DB 1; Length 457;	
Best Local Similarity	52.8%;	Pred. No. 6,1e-29;	
Matches 104; Conservative 32;		Mismatches 55; Indels 6; Gaps 5;	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
FT DOMAIN 318 374		IG-LIKE C2-TYPE 3	
FT CARBOHYD 187 187		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 298 298		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 323 323		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 392 392		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT DISULFID 42 112			
FT DISULFID 159 188			
FT DISULFID 328 370			
FT LIPID 418 418		S-palmitoyl cysteine (By similarity).	
FT LIPID 421 421		S-palmitoyl cysteine (By similarity).	
FT VARSPDIC 1 240		Missing (in isoform 2).	
FT VARSPDIC 1 240		/FTID=VSP 002489.	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
Query Match	36.1%;	Score 475; DB 1; Length 457;	
Best Local Similarity	52.8%;	Pred. No. 6,1e-29;	
Matches 104; Conservative 32;		Mismatches 55; Indels 6; Gaps 5;	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
FT DOMAIN 318 374		IG-LIKE C2-TYPE 3	
FT CARBOHYD 187 187		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 298 298		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 323 323		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 392 392		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT DISULFID 42 112			
FT DISULFID 159 188			
FT DISULFID 328 370			
FT LIPID 418 418		S-palmitoyl cysteine (By similarity).	
FT LIPID 421 421		S-palmitoyl cysteine (By similarity).	
FT VARSPDIC 1 240		Missing (in isoform 2).	
FT VARSPDIC 1 240		/FTID=VSP 002489.	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
Query Match	36.1%;	Score 475; DB 1; Length 457;	
Best Local Similarity	52.8%;	Pred. No. 6,1e-29;	
Matches 104; Conservative 32;		Mismatches 55; Indels 6; Gaps 5;	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
FT DOMAIN 318 374		IG-LIKE C2-TYPE 3	
FT CARBOHYD 187 187		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 298 298		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 323 323		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 392 392		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT DISULFID 42 112			
FT DISULFID 159 188			
FT DISULFID 328 370			
FT LIPID 418 418		S-palmitoyl cysteine (By similarity).	
FT LIPID 421 421		S-palmitoyl cysteine (By similarity).	
FT VARSPDIC 1 240		Missing (in isoform 2).	
FT VARSPDIC 1 240		/FTID=VSP 002489.	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
Query Match	36.1%;	Score 475; DB 1; Length 457;	
Best Local Similarity	52.8%;	Pred. No. 6,1e-29;	
Matches 104; Conservative 32;		Mismatches 55; Indels 6; Gaps 5;	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
FT DOMAIN 318 374		IG-LIKE C2-TYPE 3	
FT CARBOHYD 187 187		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 298 298		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 323 323		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD 392 392		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT DISULFID 42 112			
FT DISULFID 159 188			
FT DISULFID 328 370			
FT LIPID 418 418		S-palmitoyl cysteine (By similarity).	
FT LIPID 421 421		S-palmitoyl cysteine (By similarity).	
FT VARSPDIC 1 240		Missing (in isoform 2).	
FT VARSPDIC 1 240		/FTID=VSP 002489.	
SEQUENCE 457 AA; 51296 MM; 181DA7527CB00F33 CRC64;			
Query Match	36.1%;	Score 475; DB 1; Length 457;	
Best Local Similarity	52.8%;	Pred. No. 6,1e-29;	
Matches 104; Conservative 32;		Mismatches 55; Indels 6; Gaps 5;	
SEQUENCE 457 AA; 51296 MM; 181DA7			

RA Brownstein M.J., Usdin T.B., Toehiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Folley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch J.E., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RN SEQUENCE OF 205-240 FROM N.A.  
 RX MEDLINE=91267564; PubMed=1711009;  
 RA Yoshikawa K., Seto M., Ueda R., Obata Y., Notake K., Yokochi T.,  
 RA Takahashi T.,  
 RT "Molecular cloning of the gene coding for the human T cell  
 RT differentiation antigen CD7.";  
 RL Immunogenetics 33:352-360(1991).  
 RN [5]  
 RN TOPOLOGY.  
 RP MEDLINE=90063052; PubMed=2479685;  
 RA Ware R.E., Searce R.M., Dietz M.A., Starmer C.F., Palker T.J.,  
 RA Haynes B.F.,  
 RT "Characterization of the surface topography and putative tertiary  
 RT structure of the human CD7 molecule.";  
 RL J. Immunol. 143:3632-3640(1989).  
 RN [6]  
 RN INTERACTION WITH SECTM1.  
 RX MEDLINE=20119303; PubMed=10652336;  
 RA Lyman S.D., Escobar S., Rousseau A.-M., Armstrong A., Fanslow W.C.,  
 RT "Identification of CD7 as a cognate of the human K12 (SECTM1)  
 RT protein.";  
 RL J. Biol. Chem. 275:3431-3437(2000).  
 CC -1- FUNCTION: Not yet known.  
 CC -1- SUBUNIT: Interacts with SECTM1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD7 entry;  
 CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd7.htm.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/  
 CC or send an email to license@ebi.ac.uk).  
 CC -----  
 DR EMBL: X06180; CAA29546.1; -;  
 DR EMBL: M37271; AAA51953.1; -;  
 DR EMBL: BC009293; AAH09293.1; -;  
 DR EMBL: BC013287; AAH13287.1; -;  
 DR EMBL: D00749; BAA00646.1; -;  
 DR EMBL: D00747; BAA00646.1; JOINED.  
 DR EMBL: D00748; BAA00646.1; JOINED.  
 DR PIR: A39016; A39016.  
 DR HSSP: P01607; IREI.  
 DR Genew; HGNC:1695; CD7.  
 DR MIM: 186820; -;  
 DR GO: GO:0016021; C:integral to membrane; TAS.  
 DR GO: GO:0005624; C:membrane fraction; TAS.  
 DR GO: GO:0005886; C:plasma membrane; TAS.  
 DR GO: GO:0004872; F:receptor activity; TAS.  
 DR GO: GO:0006955; P:immune response; TAS.  
 DR GO: GO:0042110; P:T-cell activation; TAS.  
 DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. .; NAS.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003599; IG.  
 DR Pfam; PF00047; Ig; 1.

DR SMART: SM00409; IG; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 KW T-cell; Signal; Immune response; Antigen; Transmembrane; Glycoprotein;  
 KW Immunoglobulin domain; Receptor; Lipoprotein; Palmitate; Repeat.  
 FT SIGNAL 1 25  
 FT CHAIN 26 240  
 FT DOMAIN 26 180  
 FT TRANSMEM 181 201  
 FT DOMAIN 202 240  
 FT DOMAIN 26 180  
 FT DOMAIN 145 180  
 FT DISULFID 35 142  
 FT DISULFID 48 114  
 FT LIPID 198 198  
 FT CARBOHYD 45 45  
 FT CARBOHYD 96 96  
 FT REPEAT 145 153  
 FT REPEAT 154 162  
 FT REPEAT 163 171  
 FT REPEAT 172 180  
 FT REPEAT 180 180  
 SQ SEQUENCE 240 AA; 25409 MW; EBBCE08279552108 CRC64;  
 Query Match 25.2%; Score 332; DB 1; Length 240;  
 Best Local Similarity 36.4%; Pred. No. 2,2e-18;  
 Matches 102; Conservative 18; Mismatches 54; Indels 106; Gaps 10;  
 QY 6 PRRHLVLTQLAL--LPATQGNKVLGK-----GDTVELCTASQKSIQFHWKNS 56  
 DB 4 PRRHLVLTQLALRGLPALAAQEVQSPHCTVVGASVITCSTSG----- 52  
 QY 57 NQKILGNGSPFLTK-GPS-----KANDRADSRSLWDQGNFLILKN 98  
 DB 53 -----GLNGILYLRQGPQPDIIYEDGVPTDRFRGRIDFSGS---QDVLITTHR 103  
 QY 99 LKIEDSDTYICEVEDKEEVQLLVFGLTANSDTHLLQGSLLTLESPPGSSPSVQCRSP 158  
 DB 104 LQSLDGTGYTC-----QALIV----- 120  
 QY 159 RGNKIQGKTLVSQLELDQSGTWCTVLQONQKVEFKIDIVPRASALPAPPGSALPDP 218  
 DB 121 ---NVGSGTL-VLVTEEGSQGNHRS-----DAPPRASALPAPPGSALPDP 164  
 QY 219 QTSALPDPAPASALPALAVISFLGLGVACVILARR 258  
 DB 165 QTSALPDPAPASALPALAVISFLGLGVACVILARR 204  
 RESULT 16  
 CAML\_RAT STANDARD; PRT; 1259 AA.  
 AC Q05695;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neural cell adhesion molecule L1 precursor (N-CAM L1).  
 GN L1CAM OR CAML1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=91372414; PubMed=1894011;  
 RA Mura M., Kobayashi M., Asou H., Dyemura K.,  
 RT "Molecular cloning of cDNA encoding the rat neural cell adhesion  
 RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced  
 RT by differential splicing.";  
 RL FEBS Lett. 289:91-95(1991).  
 CC -1- FUNCTION: Cell adhesion molecule with an important role in the  
 CC development of the nervous system. Involved in neuron-neuron  
 CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds  
 CC to axonin on neurons.

```

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q05695-1; Sequence=Displayed;
CC Name=2; Synonym=ALics;
CC IsoId=Q05695-2; Sequence=VSP_002592;
CC TISSUE SPECIFICITY: Isoform 2 is predominantly found in the brain,
CC while isoform 1 is found in the peripheral nervous system.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NG2CAM family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59149; CA41860.1; -.
CC PIR: S36126; S36126.
CC HSSP: P20241; ICFB.
CC InterPro: IPR008957; FN_III-like.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR007110; IG-like.
CC InterPro: IPR003598; IG_c2.
CC Pfam: PF00041; fn3; 4.
CC Pfam: PF00047; Ig; 6.
CC SMART: SM00060; FN3; 4.
CC SMART: SM00408; IGC2; 5.
CC PROSITE: PS00835; IG_LIKE; 6.
CC Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
CC Transmembrane; Repeat; Immunoglobulin domain; Signal;
CC Alternative splicing.
CC FT SIGNAL 1 19 BY SIMILARITY.
CC CHAIN 20 1259 NEURAL CELL ADHESION MOLECULE L1.
CC DOMAIN 20 1122 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1123 1145 POTENTIAL.
CC DOMAIN 1146 1259 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 35 128 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 138 225 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 239 327 IG-LIKE C2-TYPE 3.
CC FT DOMAIN 332 419 IG-LIKE C2-TYPE 4.
CC FT DOMAIN 424 506 IG-LIKE C2-TYPE 5.
CC FT DOMAIN 517 600 IG-LIKE C2-TYPE 6.
CC FT DOMAIN 827 896 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 932 994 FIBRONECTIN TYPE-III 2.
CC DOMAIN 1032 1093 FIBRONECTIN TYPE-III 3.
CC SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
CC SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
CC FT DISULFID 57 113 BY SIMILARITY.
CC FT DISULFID 157 208 BY SIMILARITY.
CC FT DISULFID 263 311 BY SIMILARITY.
CC FT DISULFID 353 403 BY SIMILARITY.
CC FT DISULFID 447 496 BY SIMILARITY.
CC FT DISULFID 538 590 BY SIMILARITY.
CC FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1021 1021 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASPLIC 1179 1182 Missing (in isoform 2).
FT FTId=VSP_002592.
SQ SEQUENCE 1259 AA; 140934 MW; 0F12A7C4415F3C08 CRC64;

Query Match 10.1%; Score 133; DB 1; Length 1259;
Best Local Similarity 24.4%; Pred. No. 0.024;
Matches 57; Conservative 31; Mismatches 96; Indels 50; Gaps 9;

OY 10 LLLVLTGLALLPAALQGNKVVYLGKKDPVYELCTASQKKSIQ--PFWNSNOKIKLNGQS 67
DB 507 ILANLVQVEATQITQPPSTIEKRGARTFTQASFPDSLASITWKGDR----- 557
OY 68 FLTRGPSKLANDRSRLMDQGNFPIIKNLKIEDSDTYIC---EVEDQKEVOLLVF 123
DB 558 -----DLQGRGSDKVFIEDG--LVYSLDVSDGDYSCVASTELDEVESRAQLLV 608
OY 124 GLTN-----SDTHLLQGOSPLTLESPPSSPSVQCRSP-----RKNITQGGKTL 169
DB 609 GSPGPVPHLELSDRHLLKQSQVHLVSW-----SPAEHNSPIEKYDIEFEDKEMAPEKMF 662
OY 170 SVSGLBLDSDGTCTVCTVQNGKKVE--FKIDIVRASALPAPPGSALPDPOTA 221
DB 663 SLGKV---PQNTSTLTKLSPVHYTRVAINKKYGPSPVSETVTPPEAA 712

RESULT 17
CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplow D., Fruh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
superfamily with binding domains similar to fibronectin.";
RL Nature 334:701-703(1988).
CC -1- FUNCTION: Cell adhesion molecule with an important role in the
development of the nervous system. Involved in neuron-neuron
adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
to axonin on neurons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NG2CAM family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X12875; CA431368.1; -.
CC PIR: S05479; S05479.
CC HSSP: P20241; ICFB.
CC MGD; MGI:96721; L1cam.

```

Query Match	Similarity	Score	DB	Length
Best Local Similarity	23.9%	Pred. No. 0.034		1260
Matches	56	Conservative	33	Mismatches 95; Indels 50; Gaps 9
Db	10	LLVLVLALLLPATATGKVKVLGGKDDYELFCTASOKSIQ--FHKMSNOIKILNGGS	67	
Db	507	ILANLQVEARNOITGPPRSALIEKKARPTFTTCQASFPDSLSASTWRDGR-----	557	
Qy	68	FLTKGPSKLNDRPADSRSLMDQGNFPLIIKNLIKIEDSDTYIC---EVEDQKEVOLLVF	123	
Db	558	-----DLQREGSDDKYFIEDGK--LVYQSADPYDQGNVCVASTELDEVSRQOLLVV	608	
Qy	124	GLTAN-----SDTHLLDQGSITLTLPSPGSSPVCQKSP-----RKNIGQSKTL	169	
Db	609	GSPGVPHLESDRLLKQSOVHLISW-----SPEDHNSPLKXYDIEFDEKNAPEKWF	662	

Oy	170	VSSQLEHDSGTWCTCYLONOKKVE--FKXIDVRSALPAPRGSALPPORCA	221
Dd	663	SLSKV-----FONGTSTLLKSPSYHYTFRYATLNKKRGSGPSVSIVTPPEAA	712
<hr/>			
RESULT_18			
ID	LAMP_HUMAN	STANDARD;	PRT; 338 AA.
AC	013439:		
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
De	Limbic system-associated membrane protein precursor (LSAMP).		
Gn	LSAMP OR LAMP.		
Os	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCl_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96235133; Pubmed=8666243;		
RA	Pimenta A.F., Fischer I., Levitt P.		
RT	"cDNA cloning and structural analysis of the human limbic-system-		
RL	associated membrane protein (LAMP)."		
RC	Gene 170:189-195(1996).		
CC	-I- FUNCTION: MEDITATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.		
CC	CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF		
CC	MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH		
CC	OF THE HYPOCAMPAL MOSSY FIBER PROTECTION (BY SIMILARITY).		
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor,		
CC	-I- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts		
CC	as well as in single layers of the superior colliculus, spinal		
CC	chord and cerebellum.		
CC	-I- SIMILARITY: Belongs to the immunoglobulin superfamily. IgION		
CC	family.		
CC	-I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U41901; AAC50569.1; -		
DR	PIR; JC4776; JG4776.		
DR	Genew; HGNC:6705; LSAMP.		
DR	MIM; 603241; -		
DR	GO; GO:0007399; P:neurogenesis; TAS.		
DR	InterPro; IPRO07110; IG-like.		
DR	InterPro; IPRO03598; IG_c2.		
DR	pfam; PF00047; ig; 3.		
DR	SMART; SMO0408; IGc2; 2.		
DR	PROSITE; PS50835; IG_LIKE; 3.		
Kw	Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;		
Kw	Repeat; Signal; Lipoprotein.		
FT	SIGNAL	1 .. 28	POTENTIAL.
FT	CHAIN	29 .. 315	LIMBIC SYSTEM-ASSOCIATED MEMBRANE PROTEIN.
FT	PROPEP	316 .. 338	REMOVED IN MATURE FORM (POTENTIAL).
FT	DOMAIN	.. 29	IG-LIKE C2-TYPE 1.
FT	DOMAIN	132 .. 214	IG-LIKE C2-TYPE 2.
FT	DOMAIN	219 .. 304	IG-LIKE C2-TYPE 3.
FT	DISULFID	53 .. 111	POTENTIAL.
FT	DISULFID	153 .. 197	POTENTIAL.
FT	DISULFID	239 .. 290	POTENTIAL.
FT	CARBOHYD	40 .. 40	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	CARBOHYD	66 .. 66	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	CARBOHYD	136 .. 136	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	CARBOHYD	148 .. 148	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	CARBOHYD	279 .. 279	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	CARBOHYD	287 .. 287	N-LINKED (GLCNAC. . ) (POTENTIAL).

```

FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 315 315 GPI-anchor amidated asparagine
FT SEQUENCE 338 AA; 37308 MW; 03455f286df5d92f CRC64;
Query Match 9.8%; Score 129; DB 1; Length 338;
Best Local Similarity 23.2%; Pred. No. 0.0095;
Matches 57; Conservative 40; Mismatches 91; Indels 58; Gaps 11;
OY 10 LLLVLTQLALPPA-----TQGNKVVAGKKGDVVELTCTASQKKSIOFHMKNSNOIKI 61
DB 14 LVLLRLCLLPLTGLPVRSDVFNRGTDNITVRQGDPAIRCVLEDDKNS-KVAMLNNSGIIF 72
OY 62 LGNGSFLTKGSPSKLNDRA--DSRSIWDQGFPLIINKIKEDSDTYICEVDQKE--- 116
DB 73 AGHD-----KMSIDPVELEKRSI--EYSIRIKVDVYDGSYTCVQTHDEKXT 121
OY 117 -EVQLVFG---LTNSDTHLQGSSTLTLESPPGSSPSVOCR--SPRGNIQGKXT- 168
DB 122 SQVYIVGVPPKISISSDVTYNESNTLYVMANGREPEVYTWHLPTGRFEGEEY 181
OY 169 LVSQLELDQSGTWTCTVLIQ-----NOKKVEFKIDIVP-----RA 203
DB 182 LEILGITRQSGKYECKAANEVSADVKQKVTVNYPPTITESKNEATGROSLKCEA 241
OY 204 SALPAP 209
DB 242 SAVPAP 247
RESULT 19
NCM2_MOUSE STANDARD; PRT; 837 AA.
ID_NCM2_MOUSE 035136; 035962;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 2 precursor (N-CAM 2) (RB-8 neural cell
adhesion molecule) (R4B12).
GN NCAM2 OR OCAM OR RNCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=BALB/c; TISSUE=Olfactory neuroepithelium;
RX MEDLINE=97368238; PubMed=9221781;
RA Yoshihara Y., Kawasaki M., Tamada A., Fujita H., Hayashi H.,
Kagamiyama H., Mori K.;
RT "OCAM: A new member of the neural cell adhesion molecule family
related to zone-to-zone projection of olfactory and vomeronasal
axons.";
RL J. Neurosci. 17:5830-5842(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=C57BL/6J; TISSUE=Olfactory epithelium;
RX MEDLINE=97476194; PubMed=9334170;
RA Alenius M., Bohm S.;
RT "Identification of a novel neural cell adhesion molecule-related gene
with a potential role in selective axonal projection.";
RL J. Biol. Chem. 272:26083-26086(1997).
CC -1- FUNCTION: May play important roles in selective fasciculation and
zone-to-zone projection of the primary olfactory axons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (long isoform) and
attached to the membrane by a GPI-anchor (short isoform).
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
OY IsoId=O35136-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=O35136-2; Sequence=VSP_002590;

```

```

CC -1- TISSUE SPECIFICITY: Expressed in subsets of both olfactory and
vomeronasal neurons in a zone-specific manner.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by, and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
-----
DR EMBL; AF001287; AAB69125.1; -
DR EMBL; AF001286; AAB69124.1; -
DR EMBL; AF016619; AAC53375.1; -
DR MGI; MGI:97282; Ncam2.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; Irb3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS00835; IG_LIKE; 5.
KW Cell adhesion; Transmembrane;
KW Immunoglobulin domain; Signal; GPI-anchor; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 837
FT DOMAIN 20 697
FT TRANSMEM 698 718
FT DOMAIN 719 837
FT DOMAIN 21 108
FT DOMAIN 113 202
FT DOMAIN 208 297
FT DOMAIN 302 396
FT DOMAIN 401 491
FT DOMAIN 482 581
FT DOMAIN 594 678
FT DISULFID 42 93
FT DISULFID 136 186
FT DISULFID 232 281
FT DISULFID 322 380
FT DISULFID 422 475
FT CARBOHYD 177 177
FT CARBOHYD 219 219
FT CARBOHYD 309 309
FT CARBOHYD 406 406
FT CARBOHYD 419 419
FT CARBOHYD 445 445
FT CARBOHYD 474 474
FT CARBOHYD 562 562
FT CARBOHYD 694 837
SQ SEQUENCE 837 AA; 93203 MW; 70473B05A2D65A5 CRC64;
Query Match 9.6%; Score 127; DB 1; Length 837;
Best Local Similarity 20.3%; Pred. No. 0.041;
Matches 48; Conservative 44; Mismatches 82; Indels 62; Gaps 10;
OY 3 RG-VPEPHLLVLTQLALPPAATQGNKVVLGKKGDVVELTCTASQKKSIOFHMKNSNOIKI 61
DB 193 RGEIDFDRIIVTVVPPALIMPPQKSPNATARGSEMTLTCASSPPPTISWPNRG--KL 250
OY 62 LGNGSFLTKGSPSKLNDRA--DSRSIWDQGFPLIINKIKEDSDTYICEV-----EDQKE 116
DB 251 IEENEKYLKQ-----SNTELTVRNININDGSGSYCKATNKAGDEQKQ 293

```

QY 117 E-VOLVLEFGLTANSDFHLLQGOS-----LTLTLESPGSSPSVQC----- 155  
 DB 294 AFIQVFEV-----QPHILQKNETTSENGHVTLYCEABEPVETIKRAIDGWFSSEG 346  
 QY 156 -RSPRG-----KNIGGKTLVSQLELDOSGTWCTVLAQ-----NOKVFEKIDIVP 202  
 DB 347 DKSPDGEIEVKGQHGRLHRIKVLKSDSGRYDEASRLGCHQRSHMLIEVAPK 402

RESULT 20  
 LAMP RAT  
 ID LAMP RAT STANDARD; PRT; 338 AA.  
 AC 062813;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Limbic system-associated membrane protein precursor (LSAMP).  
 GN LSAMP OR LAMP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN 1;  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-49.  
 RC TISSUE=Hippocampus;  
 RX MEDLINE=95374785; PubMed=7646886;  
 RA Pimenta A.F., Zukareva V., Barbe M.F., Reinoso B.S., Grimley C.,  
 RA Henzel W., Fischer I., Levitt P.;  
 RT "The limbic system-associated membrane protein is an Ig superfamily  
 member that mediates selective neuronal growth and axon targeting";  
 RL Neuron 15:287-297(1995).  
 CC -1- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.  
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF  
 CC NATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH  
 CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-C-  
 CC associated cortical and subcortical regions that function in  
 CC cognition, emotion, memory, and learning.  
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS  
 CC DETECTED IN PRESUMPTIVE CORTEX, MEDIAL LIMBIC AREAS OF THE  
 CC THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN  
 CC HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGALA AND MEDIAL THALAMIC  
 CC REGION.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. Iglon  
 CC family.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; U15554; AAA86120.1; -;  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00408; IgC2; 2.  
 DR PROSITE; PSS0835; IG-LIKE; 3.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 KW Repeat; Signal; Lipoprotein.  
 FT FT SIGNAL 1 28  
 FT CHAIN 29 315  
 FT PROPEP 316 338 LIMBIC SYSTEM-ASSOCIATED MEMBRANE  
 FT DOMAIN 29 122 PROTEIN.  
 FT DOMAIN 132 214 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 219 304 IG-LIKE C2-TYPE 1.  
 FT DISULFID 53 111 IG-LIKE C2-TYPE 3.  
 FT DISULFID 153 197 POTENTIAL.  
 FT POTENTIAL.

FT DISULFID 239 290 POTENTIAL.  
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 315 315 GPI-anchor amidated asparagine  
 (Potential).  
 SQ SEQUENCE 338 AA; 37324 MW; 0B76AFDD68A39B86 CRC64;  
 Query Match 9.6%; Score 126; DB 1; Length 338;  
 Best Local Similarity 22.8%; Pred. No. 0.016;  
 Matches 56; Conservative 41; Mismatches 91; Indels 58; Gaps 11;  
 QY 10 LLLVLTALLPAA-----TGANKVVLGKKGDVLELTCTASQKSIQPHWNSNQIKI 61  
 DB 14 LVILRLCLLPGLPVRSDVFNKGTNTVRGQDTALILCVEDKIS-KVAMLNRSGLIF 72  
 QY 62 LGNGSFLTKGSPKLNDR--DSRSIMDQGNPILIKLKIEDSTYICEVDQKE--- 116  
 DB 73 AGHD-----KMSLDPRVLEKRLH-----EYSLRIQKVDYDEGSYCSYQTOHEPXT 121  
 QY 117 -EYQLVIFG-----LTNSDTHLLQGOSLTLTLESPGSSPSVQC--SPRGKNIGGKT- 168  
 DB 122 SQYLLLVQVPPKISNISDVTVNKGSNVTLVCMANGRPVITWRHLTPGREFESEEY 181  
 QY 169 LSVSLELDOSGTWCTVLAQ-----NOKVFEKIDIVP-----RA 203  
 DB 182 LEILGTRGSGKVECKANEVSADVKQVYVNPPIITTSKSNATTTGRASLKCA 241  
 QY 204 SALPAP 209  
 DB 242 SAVPAP 247

RESULT 21  
 LAMP RAT  
 ID LAMP RAT STANDARD; PRT; 739 AA.  
 AC P29534;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).  
 GN VCAM1 OR VCAM-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=92181437; PubMed=1371918;  
 RA Heselin C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,  
 RA Barkly L., Miyake K., Kincaid P., Lobb R.;  
 RT "Cloning of murine and rat vascular cell adhesion molecule-1";  
 RL Biochem. Biophys. Res. Commun. 183:163-169(1992).  
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
 CC INTEGRIN VIA4 ON LEUKOCYTES. AND MEDIATES BOTH ADHESION AND SIGNAL  
 CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A  
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE  
 CC EMIGRATION TO SITES OF INFLAMMATION.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
 CC well as on macrophage-like and dendritic cell types in both normal  
 CC and inflamed tissue.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

OC Mammalia,Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RX [1]
RX SEQUENCE FROM N.A., AND SEQUENCE OF 32-62.
RX MEDLINE=99175207; PubMed=10075727;
RA Funatsu N., Miyata S., Kumanogoh H., Shigeta M., Hamada K., Endo Y.,
RA Sokawa Y., Mekawa S.;
RT "Characterization of a novel rat brain glycosylphosphatidylinositol-
RT anchored protein (Kilon), a member of the Iglon cell adhesion molecule
RT family.";
RL J. Biol. Chem. 274:8224-8230(1999).
CC -I- FUNCTION: CELL-ADHESION (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -I- TISSUE SPECIFICITY: Highly expressed in brain.
CC -I- PTM: Glycosylated.
CC -I- SIMILARITY: Belongs to the immunoglobulin superfamily. Iglon
CC family.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB017139; BAA75649.1; -.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003598; Ig_C2.
CC Pfam; PF00047; Ig; 3.
CC SMART; SM00408; IgC2; 2.
CC PROSITE; PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
CC Repeat: Signal.
KW CHAIN.
FT 1 31
FT PROPER ? 348
FT DOMAIN 32 128
FT DOMAIN 133 215
FT DOMAIN 219 307
FT DISULFID 54 112
FT DISULFID 154 197
FT DISULFID 239 291
FT CARBOHYD 67 67
FT CARBOHYD 149 149
FT CARBOHYD 269 269
FT CARBOHYD 280 280
FT CARBOHYD 288 288
FT CARBOHYD 301 301
FT SEQUENCE 348 AA; 37858 MW; 37E90D1C7D24ACAB CRC64;
-----
Query Match 9.3%; Score 122; DB 1; Length 348;
Beet Local Similarity 23.0%; Pred. No. 0.033;
Matches 53; Conservative 35; Mismatches 94; Indels 48; Gaps 10;
-----
OY PAATGGKVVVIGKKGVETLTCTAAGKSIQFMKMSNOIKILGNOSFLLTGP---SK 75
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 35 PMAAVDMVLV--RKGDPAVLRCTLEDAS--KGMALNRSSIIIFAG--GDKMSVDPRAVIST 89
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 76 LNDRAADSRSLMDQGNFLLIKKLIKEDSDTYICEVEDQKEVQLVFGTLA----- 127
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 90 LNRK-----DYSLQIQNVVDVDDPGPYCSVQTQHTPTMQVH--LTQVVPKTYD 137
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 128 -NSTHTLLOQSLLTLTIESPPSSPSVQCR--SPRGKNIQGGTISVSQLELDSGWTWC 164
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 138 ISNDMTINTEGNTVLTLCATGKEPALSMWHISPSAKPFENGOYLDIYGTTRDOAGBYEC 197
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 185 TV-----LQNGKKVEFKIDIVPRASALPAPT-----GSAALPDP 218
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 198 SAENDVSPFDVKKVVVNVNPAFPIQTLEKSGTVPRGSGLRCEGAGVPPP 247
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

ID	NAME	STANDARD	PRT	349 AA
LN	LACH_SCHAM			
AC	Q26474			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Lachesin precursor.			
GN	LAC.			
OS	Schistocerca americana (American grasshopper).			
OC	Eumetazoa; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;			
OC	Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridoidea;			
OC	Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.			
OX	NCBI_Taxid=7009;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=9403693; PubMed=8223276;			
RA	Karlstrom R.O., Wilder L.P., Bastiani M.J.;			
RT	"Lachesin: an immunoglobulin superfamily protein whose expression			
RL	correlates with neurogenesis in grasshopper embryos.";			
CC	Development 118:509-522(1993).			
CC	-1- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND			
CC	AXON OUTGROWTH.			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED BY ALL NEUROGENIC CELLS EARLY, BUT			
CC	ONLY THOSE CELLS THAT BECOME NEUROBLASTS CONTINUE TO EXPRESS IT.			
CC	EXPRESSED BY NEUROBLASTS, GANGLION MOTHER CELLS AND NEURONS EARLY			
CC	IN THEIR LIVES, BUT EXPRESSION BECOMES RESTRICTED TO A SUBSET OF			
CC	NEURONS AS DEVELOPMENT PROGRESSES. EXPRESSED BY SENSORY NEURONS AS			
CC	THEY DELAMINATE FROM THE BODY WALL ECTODERM. IT IS ALSO PRESENT ON			
CC	GROWING AXONS OF THE CNS AND BECOMES RESTRICTED TO A			
CC	SUBSET OF AXONS LATER IN DEVELOPMENT.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS			
CC	FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL			
CC	NERVOUS SYSTEMS.			
CC	-1- PTM: The N-terminus is blocked.			
CC	-1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.			
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; L13256; AAC37185.1; .			
DR	HSSP; P80362; IMTL.			
DR	InterPro; IPR007110; IG_1-like.			
DR	InterPro; IPR003598; IG_C2.			
DR	Pfam; PF00047; Ig_3.			
DR	SMART; SM00408; IGC2; 2.			
DR	PROSITE; PS00835; IG_LIKE; 3.			
RW	Immunoglobulin domain; Cell			
KW	Repeat; Signal; Lipoprotein.			
FT	SIGNAL 1 18			
FT	CHAIN 19 332			
FT	PROPEP 333 349			
FT	DOMAIN 22 127			
FT	DOMAIN 132 218			
FT	DOMAIN 222 315			
FT	DISULFID 43 110			
FT	DISULFID 154 201			
FT	DISULFID 244 299			
FT	CARBOHYD 137 137			
FT	LIPID 332 332			
SO	SEQUENCE 349 AA; 38974 MW; 5F139A44BF849689 CRC64;			
Query Match:	9.1%;	Score 120;	DB 1;	Length 349;
Best Local Similarity:	25.3%;	Pred. NO. 0.04;		
Matches 58;	Conservative 35;	Mismatches 100;	Indels 36;	Gaps 10;

```

OY 34 GDVLELCTSSOKKSLQIFHH-----KNSNQIKIINGQSSFLTKPSPSKINDPADSRSLMDQ 89
Dy 36 GGVELECSQVQADYVYVMKMKYDRRQVDPPLISTSSSLINDSRALALYDPASS----- 91
OY 90 GNPPLIITKLIKIEDSDTYICEV-----EDQKEVQLVFC--LITANSPTHLLOQSGSLT 141
Db 92 -TYTLQIKDIQIEDDAGFYQCOVITGLNNKITAEDVLDVLRPPVYISDSTSLVSEGOAV 150
OY 142 TLSEPPSS--SPSYQCSS-----PRKINIOGGTKTSSVSOLEDSGTWTCYVLON---- 169
Db 151 RLECYAGGYAPKVSWMRENNALLPTGGSITYRGVNLKISIGKEDRGTYTC-VAENGVK 209
OY 190 --OKRYEFKIDIVPRASALPAPPTGSALP---DPQTASALDPDPPASAL 233
Db 210 GARINIAVEVEF--PPVITVPRPRILGQLQYDMDLCEHVEAYPPPAITWML 257

RESULT 24:
CAML_FUGRU STANDARD; PRT; 1277 AA.
AC O98902;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule LI precursor (N-CAM LI) (LI-CAM).
LN L1CAM.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
CN NCBL_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain, and Muscle;
RX MEDLINE=98147998; Pubmed=9479034;
RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,
RA Platzner M., Drecher B., Joutel M., Kennrick S., Rosenthal A.;
RT "The neural cell adhesion molecule LI: genomic organisation and
RT differential splicing is conserved between man and the pufferfish
RT Fugu."
RL Gene 208:7-15 (1998).
CC -I- FUNCTION: Cell adhesion molecule with an important role in the
CC development of the nervous system. Involved in neuron-neuron
CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
CC to axonin on neurons.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Brain;
CC IsoId=G98902-1; Sequence=Displayed;
CC Name=2; Synonyms=Muscle;
CC IsoId=G98902-2; Sequence=VSP_050474, VSP_050475;
CC -I- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC LI/neurofascin/NG2CAM family.
CC -I- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```

```

DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 5.
DR SMART; SM00408; IGc2; 3.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW Immunoglobulin domain; Signal; Alternative splicing.
FT SIGNAL 1 34
FT CHAIN 35 1277
FT DOMAIN 30 1135
FT TRANSMEM 1136 1156
FT DOMAIN 1157 1277
FT DOMAIN 51 140
FT DOMAIN 150 241
FT DOMAIN 256 344
FT DOMAIN 349 437
FT DOMAIN 443 528
FT DOMAIN 532 623
FT DOMAIN 628 726
FT DOMAIN 727 825
FT DOMAIN 826 932
FT DOMAIN 933 1028
FT DOMAIN 1029 1122
FT DISULFID 72 129
FT DISULFID 173 224
FT DISULFID 280 328
FT DISULFID 370 421
FT DISULFID 465 514
FT DISULFID 554 607
FT CARBOHYD 317 503
FT CARBOHYD 520 520
FT CARBOHYD 531 531
FT CARBOHYD 794 794
FT CARBOHYD 839 839
FT CARBOHYD 1035 1035
FT CARBOHYD 1046 1046
FT CARBOHYD 1068 1068
FT CARBOHYD 1083 1083
FT CARBOHYD 1108 1108
FT VASAPLIC 42 47
FT VASAPLIC 1190 1193
FT SEQUENCE 1277 AA; 141954 MW; 284BB49BA9A42C27 CRC64;
Query Match 9.1%; Score 120; DB 1; Length 1277;
Best Local Similarity 24.7%; Pred. No. 0.23;
Matches 55; Conservative 34; Mismatches 92; Indels 42; Gaps 9;
QY 34 GDTVELTCTASQKSIQFHWKNSQIKILNGQSFLLTKGPKLNDRADRSRLMDQGNFP 93
Db 363 GETVLDLDOAGDGISSPRTTW-----TVNGVPLSATSLPRSLTRESSG-- 405
QY 94 LIINKLKEDSDTYICEVEDOKER-----VQLVFGITANSDFH-LLOGSLTL-- 141
Db 406 LILKDVIFGDTAIVQCOQASNRGHTLLANTNVYIELPPQILTEGNGTYTFEGGKALLEC 465
QY 142 -TLRSP-----PSSPVSQCRSPKGNKIOGKTLVSQLELODSGTCTVLONOKVE 194
Db 466 ETFSPPKRYTWESSISILLADPRVNLITNG-GLEINAVSHDEBGIYTCLVQSSNITSVN 524
QY 195 FKIDIVPPASALPAPPTGASALPQPTA---SALPDPAPASAL 233
Db 525 AEVEVLRNTVIL-SPPQALRLQPKTALFTCLVYTDPLSPL 566
RESULT 25
CONT_CHICK STANDARD; PRT; 1010 AA.
AC P14781; P10450;
DT 01-MAR-1989 (Rel. 10, Created)

```

```

DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DR Contactin precursor (Neural cell recognition molecule F11).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90180453; PubMed=2627374;
RA Brummenndorf T., Wolff J.M., Ratner F., Rathjen F.G.;
RT "Neural cell recognition molecule F11: homology with fibronectin type
RL Iii and immunoglobulin type C domains.",
RN Neuron 2:1351-1361(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=White Leghorn;
RC MEDLINE=8908597; PubMed=3049624;
RA Ranscht B., Dours M.T.;
RT "Sequence of contactin, a 130-kD glycoprotein concentrated in areas
RT of interneuronal contact, defines a new member of the immunoglobulin
RT supergene family in the nervous system.";
RL J. Cell Biol. 107:1561-1573(1988).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=89286606; PubMed=2735929;
RA Wolff J.M., Brummenndorf T., Rathjen F.G.;
RT "Neural cell recognition molecule F11: membrane interaction by
RL covalently attached phosphatidylinositol.",
RX Biochem. Biophys. Res. Commun. 161:931-938(1989).
RL -1- FUNCTION: Mediates cell surface interactions during nervous system
development.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 4 fibronectin type Iii domains.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to
CC framehifts in positions 1002 and 1007.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X14877; CA33018.1; -.
DR EMBL; Y00813; CA68753.1; ALT_FRAME.
DR PIR; S01998; S01998.
DR HSSP; P40189; 1B0U.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00047; Ig; 6.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain; Membrane; Glycoprotein; Lipoprotein; Signal;
KW GPI-anchor; Cell adhesion; Repeat.
FT SIGNAL 1 19
FT CHAIN 20 984
FT PROPEP 985 1010
FT DOMAIN 123 123
FT DOMAIN 132 215
FT DOMAIN 232 317
FT DOMAIN 322 398
FT DOMAIN 404 491
FT DOMAIN 496 592
FT DOMAIN 593 599
FT DOMAIN 600 701
CONTRACTIN.
REMOVED IN NATURE FORM (POTENTIAL).
IG-LIKE C2-TYPE 1.
IG-LIKE C2-TYPE 2.
IG-LIKE C2-TYPE 3.
IG-LIKE C2-TYPE 4.
IG-LIKE C2-TYPE 5.
IG-LIKE C2-TYPE 6.
GLY/PRO-RICH.
FIBRONECTIN TYPE-III 1.

```

Query Match	Best Local	Similarity	Score	DB 1	Length	1010
Matches	51	Conservative	33	Mismatches	96	Indels
					68	Gaps
QY	14	LQALLPATQGN--KVLGKGDTVELTCTASQKSKIOPHMKNSNOIKLNGQSF	70			
DB	397	LKIIVASPTPELNPMPKKKILIAKGGRVILIECKPKAPKPKFSWSKGTLLVNG	449			
QY	71	KPGSLNRADRSRRSLMDQGNPFLIKLKIIEDSTYICVENDQKEEVQLLVGLTANS	130			
DB	450	-----SRHIWDDGSLLEI--VNTKLDBGRYTCFAPENNRK-----ANST	487			
QY	131	THLGGSGSLTTLT-----ESPFGSSPSVQC-----RS	157			
DB	488	GVLEMTENTRITLAVLNDVYGEANATQCLASHDPTLDLFTIWSLNGFVDFEKEHNY	547			
QY	158	PRGNKIQGKTLVSQLELDQSGTCTVLQNGKVEFKIDIVPRASALPAPPTGSALPD	217			
DB	548	ERNWIKKNGSELLIKNVQLRHAGRYCTAQTIVDSSASADLVVGP--PGPGGIRIEE	605			
QY	218	-PPTASL 224				
DB	606	IRDTAVAL 613				
RESULT 26						
CAM1_BRARE		STANDARD;	PRT, 1197 AA.			
ID	CAM1_BRARE					
AC	Q90478;					
DT	10-OCT-2003 (Rel. 42, Created)					
DT	10-OCT-2003 (Rel. 42, Last sequence update)					
DT	15-MAR-2004 (Rel. 43, Last annotation update)					
DE	Neural cell adhesion molecule L1.1 (N-CAM L1.1) (Fragment).					
GN	NADL1.1.					
OS	Brachydanio rerio (zebrafish) (Danio rerio).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Actinopterygii; Neopeterygii; Teleostei; Ostariophysi; Cypriniformes;					
OC	Cyprinidae; Danio.					
OX	NCBI_Taxid=7955;					
RP	SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.					
RP	TISSUE=Embryo;					
RC	MEDLINE=96155762; PubMed=8568941;					
RX	Tongiorigi E., Bernhardt R.R., Schachner M.,					
RA	"Zebrafish neurons express two L1-related molecules during early					
RT	axogenesis."					
RL	J. Neurosci. Res. 42:547-561(1995).					
CC	-!- FUNCTION: Cell adhesion molecule with an important role in the					
CC	development of the nervous system. Involved in neuron-neuron					
CC	adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds					
CC	to axonin on neurons (By similarity).					

Query Match	9.0%	Score 118	DB 1	Length 1197
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- TISSUE SPECIFICITY: Expressed in postmitotic neurons in 16-36			
CC	hour embryos, including those in the brain, cranial ganglia and			
CC	otic and olfactory placodes, and in all classes of spinal			
CC	neurons.			
CC	-1- DEVELOPMENTAL STAGE: Onset of expression correlates with the			
CC	initiation of axonogenesis in 16-36 hour embryos.			
CC	-1- SIMILARITY: Belongs to the immunoglobulin superfamily.			
CC	11/neurofascin/NgCAM family.			
CC	-1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.			
CC	-1- SIMILARITY: Contains 5 fibronectin type III domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X89204; CAA61490.1; -.			
DR	PIR; T30581; T30581.			
DR	HSSP; P20241; ICFB.			
DR	ZFIN; ZDB-GENE-980526-512; nad11.1.			
DR	InterPro; IPR008957; FN_III-like.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR003962; FNIII_subd.			
DR	InterPro; IPR007110; IG-like.			
DR	InterPro; IPR003598; IG_c2.			
DR	Pfam; PF00041; fn3; 5.			
DR	Pfam; PF00047; ig; 6.			
DR	PRINTS; PR00014; FNTYPEIII.			
DR	SMART; SM0060; FN3; 5.			
DR	SMART; SM00408; IGc2; 4.			
DR	PROSITE; PS50835; IG_LIKE; 6.			
KM	Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;			
KW	Transmembrane; Repeat; Immunoglobulin domain.			
FT	NON_TER 1			
FT	DOMAIN <1 1054			
FT	TRANSSEM 1055 1075			
FT	DOMAIN 1076 1197			
FT	DOMAIN <1 58			
FT	DOMAIN 69 160			
FT	DOMAIN 165 263			
FT	DOMAIN 268 355			
FT	DOMAIN 360 442			
FT	DOMAIN 451 541			
FT	DOMAIN 546 638			
FT	DOMAIN 645 739			
FT	DOMAIN 744 849			
FT	DOMAIN 850 948			
FT	DOMAIN 952 1029			
FT	DISULFID 92 143			
FT	DISULFID 199 247			
FT	DISULFID 289 339			
FT	DISULFID 383 432			
FT	DISULFID 472 525			
FT	CARBOHYD 135 135			
FT	CARBOHYD 149 149			
FT	CARBOHYD 221 221			
FT	CARBOHYD 298 298			
FT	CARBOHYD 414 414			
FT	CARBOHYD 421 421			
FT	CARBOHYD 438 438			
FT	CARBOHYD 449 449			
FT	CARBOHYD 708 708			
FT	CARBOHYD 959 959			
FT	CARBOHYD 968 968			
FT	CARBOHYD 1002 1002			
FT	CARBOHYD 1027 1027			
SEQ	SEQUENCE 1197 AA; 132860 MW; 7CE1505EEBAC7B28 CRC64;			

Best Local Similarity 19.5%; Pred. No. 0.31;  
Matches 44; Conservative 44; Mismatches 86; Indels 52; Gaps 8;

QY 9 HLLVLQALLPAATQGNKVLGKGGDVELTCTASQKKSIOFMKNSNQIKIINGNSF 68  
DB 259 HYTVTEA--PYWTRSEEHLYAPGETVRLDCKADGIPAPNTW-SINGPVSGTD--- 312  
QY 69 LTKGPKLNDRADRSRLMDQGNFPLIKLKIEDSDTYICEVEDQKEVOLVFGILAN 128  
DB 313 -----VDRRRV---SSGLILSNVFPSTANYQCEAVNKGSIIL-----N 351  
QY 129 SDTHLQGSJLTJ-----LESPPGSSPSVQCRS-----PRGNIQGK 167  
DB 352 THAVHVELPAQILTPDERLYQATAGQVWLDCRFGSPFKIHWEILDISIPALSNKISQ 411  
QY 168 T-----LSVSQLDQDSGTCTVCLONQKVEFKDIVRASALPAP 209  
DB 412 TTNGSLKISNVEBDSNRYTCSVSETNKSISADVEVLRKIVGPP 457

## RESULT 27

OPCM\_BOVIN STANDARD; PRT; 345 AA.  
AC P11834;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Opioid binding protein/cell adhesion molecule precursor (OBCAM)  
DE (Opioid-binding cell adhesion molecule) (OPCM).  
GN OPCML OR OBCAM OR OCAM.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;  
RP [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=89251576; PubMed=2721489;  
RA Schofield P.R., McFarland K.C., Hayflick J.S., Wilcox J.N.,  
RA Cho T.M., Roy S., Lee N.M., Loh H.H., Seeburg P.H.;  
RT "Molecular characterization of a new immunoglobulin superfamily  
RT protein with potential roles in opioid binding and cell contact."  
RL EMBL J. 8:489-495 (1989).  
CC -1- FUNCTION: Binds opioid in the presence of acidic lipid; probably  
CC involved in cell contact.  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
CC similarity).  
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON  
CC family.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC  
CC EMBL: X12672; CAA31192.1; .  
DR PIR: S03199; S03199.  
DR InterPro: IPR007110; IG\_1like.  
DR InterPro: IPR003598; IG\_c2.  
DR Pfam: PF00047; IG\_3.  
DR SMART: SM00408; IGc2; 2.  
DR PROSITE: PS50835; IG\_LIKE; 3.  
KM Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
KW Repeat; Signal; Lipoprotein.  
FT SIGNAL 1 27  
FT CHAIN 28 322  
FT PROPER 323 345 OPIOID BINDING PROTEIN/CELL ADHESION  
FT MOLECULE.  
FT REMOVED IN MATURE FORM (POTENTIAL).

FT DOMAIN 39 126 IG-LIKE C2-TYPE 1.  
FT DOMAIN 136 219 IG-LIKE C2-TYPE 2.  
FT DOMAIN 223 310 IG-LIKE C2-TYPE 3.  
FT DISULFID 57 115 POTENTIAL.  
FT DISULFID 157 202 POTENTIAL.  
FT DISULFID 244 296 POTENTIAL.  
FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT LIPID 322 322 GPI-anchor amidated asparagine  
FT (Potential).  
SQ SEQUENCE 345 AA; 37914 MW; DIECC8D5E7BDCB19 CRC64;

Query Match 8.9%; Score 117.5; DB 1; Length 345;  
Best Local Similarity 22.4%; Pred. No. 0.073;  
Matches 59; Conservative 43; Mismatches 100; Indels 61; Gaps 10;

QY 11 LVLVLQALLPAATQGNKVLG-----KKGDVLELTCTASQKKSIOFMKNSN 57  
DB 14 LVVSLRLPLVPTGVPRSGDATFPKAMNVTYVQGESATLRTCTIDRVT-RVAMLRNS 72  
QY 58 QIKILGN-----QGSFLTKGPKLNDRADRSRLMDQGNFPLIKLKIEDSDTYICE 110  
DB 73 TILVANGDKSIDPRVILVNTPTQ-----YSIMIQNVVDYDGGPYTCS 116  
QY 111 VE-----DQKEVOLVFG-----LTNSDTHLQGSJLTJLESPPGSSPSVQCRSPGKN 162  
DB 117 VQTNHFKTSRVHLIVQVPOIMNISDVTVNEGSSVTLCLAIQGRDEPTVWHLISKE 176  
QY 163 IQG-----GKLSVSQLDQDSGTCTVCLQ-----NOKYVEFKDIVRASALPAPRTGS 213  
DB 177 GQGFSEDEYLEIISDIRDSGEYECALNDVAPDVAKYKIVNYPYIS--RAKQTV 234  
QY 214 ALPPQTSALPDPASALPAA 236  
DB 235 SVGKGILSC-----EASAVPMA 252

## RESULT 28

LAMP\_CHICK STANDARD; PRT; 338 AA.  
AC Q98919;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Limbic system-associated membrane protein precursor (E19S) (CHLAMP,  
DE G19-isoform). (Chicken).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
NCBI\_TaxID=9031;  
RP [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97157768; PubMed=9004047;  
RA Wilson D.J.A., Kim D.S., Clarke G.A., Marshall-Clarke S., Moss D.J.;  
RT "A family of glycoproteins (GP55), which inhibit neurite outgrowth,  
RT are members of the Ig superfamily and are related to OBCAM,  
RT neurotrophin, LAMP and CEPU-1."  
RL J. Cell Sci. 109:3129-3138 (1996).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97358596; PubMed=9215692;  
RA Brumendorf T., Spaltmann F., Treubert U.;  
RT "Cloning and characterization of a neural cell recognition molecule  
RT on axons of the retinorecortical system and spinal cord."  
RL Eur. J. Neurosci. 9:1105-1116 (1997).  
CC -1- FUNCTION: Mediates selective neuronal growth and axon targeting.

Query Match	Best Local Match	Similarity	Score	DB 1	Length	338
Matches	55	Conservative	41	Mismatches	93	Indels
					56	Gaps
					11	
10	LLVLVQLQLLPPA	-----	TOGKVVILGKKGDVETVETCTASQKKSIOFMKNSNQIKI	61		
14	LVLLRLCLLPTGLPEVRSDVDFTRGNDITVAGQDAIILRCFEDSS	-----	KVAMLNRSGLIF	72		
62	LGNQGSPLTKGSPSKINDRAD	-----	SRSLMDQGNFPLIIKMLKIEDSTPIYCEVEDQ	116		
73	AG-----	-----	EDKMSLDPRVELEKRSPL	122		
117	EVQLVLPFG-----	-----	LTANSDTHLLQGGSLLTLESPPGSSPSVQCR	169		
123	QVYLIVQVPPKISNISSDITVNEGSNTLVLCMANGRPBIVITWRHL	-----	TPGKEFGSEBYL	182		
170	SVSOLDELQSGTWCTVLO	-----	NQKVEFKIDIVP	204		
183	EILGTRQSGKYECKKANEVASAVYKQVRYVTVNPPITTESKSNKATGQALLRCBAS	-----	242			
205	ALPAP	209				
243	AVPTP	247				

ID	V_CAM1_HUMAN	STANDARD;	PTT;	739 AA.
AC	P19320;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen) (INCAM-100).			
CN	VCAM1 OR L1CAM.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Umbilical vein;			
RX	MEDLINE=91016951; PubMed=1699207;			
RA	Polte T., Newman W., Gopal T.V.;			
RT	"Full length vascular cell adhesion molecule 1 (VCAM-1).";			
RL	Nucleic Acids Res. 18:5901-5901(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90090619; PubMed=2688899;			
RA	Osborn L., Hession C., Tizard R., Vassallo C., Luhowskyj S.,			
RA	Chi-Rosso G., Lobb R.;			
RT	"Direct expression cloning of vascular cell adhesion molecule 1, a			
RT	cytokine-induced endothelial protein that binds to lymphocytes.";			
RL	Cell 59:1203-1211(1989).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91352090; PubMed=1715583;			
RA	Cybulsky M.I., Fries J.W.U., Williams A.J., Sultan P., Eddy R.,			
RA	Byers M., Shows T., Glimbert M.A. Jr., Collins T.;			
RT	"Gene structure, chromosomal location, and basis for alternative mRNA			
RT	splicing of the human VCAM1 gene.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863(1991).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91201302; PubMed=1707873;			
RA	Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,			
RA	Chi-Rosso G., Luhowskyj S., Lobb R., Osborn L.;			
RT	"Cloning of an alternate form of vascular cell adhesion molecule-1			
RT	(VCAM1).";			
RL	J. Biol. Chem. 266:6682-6685(1991).			
RN	[5]			
RP	SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND			
RP	LEU-716.			
RA	Rider M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,			
RA	Rajkumar N.R., Toth E.J., Yi O., Nickerson D.A.;			
RL	Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SEQUENCE FROM N.A. (ISOFORM LONG).			
RC	TISSUE=Retinal pigment epithelium;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,			
RA	Altshul S.F., Zeeberg B., Bielow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Ditachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,			
RA	Brownstein M.J., Uedin T.B., Toehiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Boesk S.A., McGowan P.J., McKernan K.J., Malek J.A., Gnatratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodegrem E.J., Lu X., Gibbs R.A.,			
RA	Faney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzyzanski M.I., Skalek U., Smailus D.E.,			
RA	Scherer A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
RX MEDLINE=95147978; PubMed=7531291;
RA Jones E.Y., Harlos K., Bottomley M.J., Robinson R.C., Driscoll P.C.,
RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.;
RT "Crystal structure of an integrin-binding fragment of vascular cell
RT adhesion molecule-1 at 1.8-A resolution.";
RL Nature 373:539-544(1995).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX MEDLINE=95296382; PubMed=7539925;
RA Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,
RA Browning B., Osborn L.;
RT "The crystal structure of an N-terminal two-domain fragment of
RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
RT interaction.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718(1995).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RA Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,
RA Osborn L.;
RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
RT resolution.";
RL Acta Crystallogr. D 52:369-379(1996).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P19320-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P19320-2; Sequence=VSP_002580;
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).
CC -1- PTM: Sialoglycoprotein.
CC -1- DISEASE: May play an important role in the genesis of
CC atherosclerosis and rheumatoid arthritis.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD106 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53051; CAA37218.1; -;
DR EMBL; M30257; AAA51917.1; ALT_TERM.
DR EMBL; M73355; AAA61270.1; -;
DR EMBL; M60335; AAA61269.1; -;
DR EMBL; AF536818; AAA61901.1; -;
DR PIR; A41288; A41288.
DR PIR; B41288; B41288.
DR PDB; 1VCA; 15-SEP-95.
DR PDB; 1VSC; 20-JUN-96.
DR PDB; 1J39; 07-NOV-01.
DR Genew; HGNC:12663; VCAM1.
DR MIM; 192225; -;
DR InterPro; IPR003987; ICAM_VCAM-1.

```

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig C2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 6.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS00835; IGC LIKE; 5.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
KW Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 739
FT DOMAIN 25 698
FT TRANSMEM 699 720
FT DOMAIN 721 739
FT DOMAIN 25 105
FT DOMAIN 109 212
FT DOMAIN 223 309
FT DOMAIN 312 399
FT DOMAIN 408 506
FT DOMAIN 511 595
FT DOMAIN 600 684
FT DISULFID 47 95
FT DISULFID 52 99
FT DISULFID 137 195
FT CARBOHYD 273 273
FT CARBOHYD 365 365
FT CARBOHYD 417 417
FT CARBOHYD 463 463
FT CARBOHYD 531 531
FT CARBOHYD 561 561
FT VASPLIC 310 402
FT FT 318 318
FT VARIANT 384 384
FT VARIANT 413 413
FT VARIANT 716 716
FT STRAND 26 30
FT STRAND 34 38
FT TURN 39 40
FT STRAND 43 50
FT STRAND 56 61
FT TURN 62 63
FT STRAND 70 74
FT TURN 75 76
FT STRAND 77 82
FT HELIX 87 89
FT STRAND 92 99
FT TURN 100 101
FT STRAND 102 114
FT STRAND 120 123
Query Match 8.9%; Score 117; DB 1; Length 739;
Best Local Similarity 22.3%; Pred. No. 0.2;
Matches 47; Conservative 38; Mismatches 84; Indels 42; Gaps 7;
Cy 8 RHLLVQLALLPAA--QGNKVVYLGKGPVVELCTASQKSIQFMKNSNQIKLIGNQ 65
Db 211 RQAVKEIQVYISPKNTYISVPSKTLQGGSVTWTCSEGLPAPEITW----- 258
Cy 66 GSFLLTKPSKLNDRADSRSLMDGNFPLIIKMLKIDSDTYICE---VEDQKEVQLT 121
Db 259 -----SKKLDNGNLQHL---SGNATLTLIAMBMEDSGIYCEGVNLGKRRKEVELI 307
Cy 122 V----FGLTANSDTHLQ--GQSITLTLBSPGSSPSVQCRSPRGKNIQ-----GKT 168
Db 308 VQEKPFVTEISPGRIAQIGDSVMLTCSYWGCESPFSFWRQTQIDSPLSGKVRSEGTNST 367
Cy 169 LSVQSLQDLSGCTWTCTVNLQNKVKVEFKIDI 199

```

Db 368 LTLSPVSENEHSLCTVTCGKKLXKGIQV 398

RESULT 30  
OPCM\_HUMAN STANDARD; PRT; 345 AA.

AC Q14982;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Opioid binding protein/cell adhesion molecule precursor (OBPCM)  
DE Opioid-binding cell adhesion molecule (OBPCM).  
GN OBPCM OR OBPCM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Occipital cortex;  
RX MEDLINE=95237612; PubMed=7721093;  
RA Shark K.B., Lee N.M.;  
RT "Cloning, sequencing and localization to chromosome 11 of a cDNA  
RT encoding a human opioid-binding cell adhesion molecule (OBPCM).";  
RL Gene 153,213-217(1995).  
CC -1- FUNCTION: Binds opioids in the presence of acidic lipids; probably  
CC involved in cell contact.  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (by  
CC similarity).  
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. Iglon  
CC family.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L34774; AAA36387.1; -.  
DR PIR: J04025;  
DR GeneW: HGNC:8143; OBPCM.  
DR MIM: 600632; -.  
DR GO: GO:0005887; C: integral to plasma membrane; TAS.  
DR GO: GO:0004985; F: opioid receptor activity; TAS.  
DR GO: GO:0007155; P: cell adhesion; TAS.  
DR GO: GO:0008038; P: neuronal cell recognition; TAS.  
DR InterPro: IPR007110; Ig-like.  
DR InterPro: IPR003598; Ig\_c2.  
DR Pfam: PF00047; Ig\_3.  
DR SMART: SM00408; IGC2; 2.  
DR PROSITE: PS50835; IG\_LIKE; 3.  
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
KW Repeat; Signal; Lipoprotein.  
FT SIGNAL 1 27  
FT CHAIN 28 322  
FT PROPEP 323 345  
FT DOMAIN 39 126  
FT DOMAIN 136 219  
FT DOMAIN 223 310  
FT DISULFID 57 115  
FT DISULFID 157 202  
FT DISULFID 244 296  
FT CARBOHYD 44 44  
FT CARBOHYD 70 70  
FT CARBOHYD 140 140  
FT CARBOHYD 285 285  
FT CARBOHYD 293 293  
FT CARBOHYD 306 306

FT LIPID 322 322 GPI-anchor amidated asparagine  
FT SEQUENCE 345 AA; 38007 MW; 87AD17BEA1AA3F4 CRC64;  
SQ

Query Match 8.8%; Score 116.5; DB 1; Length 345;  
Best Local Similarity 22.4%; Pred. No. 0.087;  
Matches 59; Conservative 43; Mismatches 100; Indels 61; Gaps 10;

QY 11 LVLQALALPAATQGNKVVLG-----KKGDVELTCTASQKSKISFHKNS 57  
DB 14 LVVVSRLFLFVPTGVPVSGDAPFPKAMDNTVTRQSGSATRTCTIDRYT-RVAMLNRS 72  
QY 58 QKILGN-----QGSFLTPGSKLDRADRSRLMDQGNPLIKNKIEDSDTYIC 110  
DB 73 TLVYAGNDKWSIDPRYIILVNTPTQ-----YSIMIQNDVYIDEGPYTS 116  
QY 111 VE-----DQEEVQLVFG---LTNSDTHLLQGSILTLSPSSSPSVQCRSPRGN 162  
DB 117 VQTDNPKTSRVHLIYQVPPQIMNISDITVNEGSSVTLICLIGRPEPTVTRHLSVKE 176  
QY 163 IQG---GKTLVSQLELDQSGTWCTVYQ-----NOKYVEKIDVPPASALPAPPTGS 213  
DB 177 GQGFVSEDEYLEISDIKRDQSGEYCSALNDVAAPDVRKVTYVNPYIS--KAKNTGV 234  
QY 214 ALPDPTASALPDPASALPAA 236  
DB 235 SVGQKGLSC-----EASAVPWA 252

RESULT 31  
LACH\_DROME STANDARD; PRT; 359 AA.

AC Q24372;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lachesin precursor.  
GN LAC.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94038693; PubMed=8223276;  
RA Karlstrom R.O., Walder L.P., Bastiani M.J.;  
RT "Lachesin: an immunoglobulin superfamily protein whose expression  
RT correlates with neurogenesis in grasshopper embryos.";  
RL Development 118:509-522(1993).  
CC -1- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND  
CC AXON OUTGROWTH.  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (by  
CC similarity).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS  
CC FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL  
CC NERVOUS SYSTEMS.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U13255; AAC37184.1; -.  
DR HSSP: P56276; 1TLK.  
DR FlyBase: FBgn0010238; Lac.  
DR InterPro: IPR007110; Ig-like.  
DR InterPro: IPR003598; Ig\_c2.



DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00408; IGc2; 2.  
 DR PROSITE; PS50835; IG LIKE; 3.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 Repeat; Signal; Lipoprotein.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 336 LACHESIN.  
 FT PROPEP 337 359 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 29 130 IG-LIKE V-TYPE.  
 FT DOMAIN 135 221 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 226 317 IG-LIKE C2-TYPE 2.  
 FT DISULFID 50 113 POTENTIAL.  
 FT DISULFID 157 204 POTENTIAL.  
 FT DISULFID 247 303 POTENTIAL.  
 FT CARBOHYD 92 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 336 336 GPI-anchor amidated alanine (Potential).  
 SQ SEQUENCE 359 AA; 39998 MW; DLF513E2B9D8666 CRC64;  
 Query Match 8.8%; Score 116.5; DB 1; Length 359;  
 Best Local Similarity 22.3%; Pred. No. 0.091; 87; Indels 43; Gaps 7;  
 Matches 47; Conservative 34; Mismatches 87; Indels 43; Gaps 7;  
 QY 34 GDTVELTCTASQKSIQFHWKNSQIKILNQSGFLTKGPKLNDRADSRSLMDQGNP 93  
 DB 43 GGTVEFDCSVOYAKEVNVLFKTDSDPVFLSTGSLVTKDSRFLRDPNNS-----TYK 97  
 QY 94 LIINKLKEDSDTYICE-----VEDQKEVOLVLFGLTANSDDLHLOQSITLTLSPFG 148  
 DB 98 LQIDIDYDTAGTYTCQVIVSTVHKVSAEVLVSRPVPISD-----NSTGVVASEG 150  
 QY 149 SSPSVQCS-----PRGNKIQGKTLVSQLELODGTWTCTVQLON- 189  
 DB 151 SEVGMECTASGYPTPTTWRRNNAILPTDSATVGNLTIRIKSVKEDRGTYIC-VADNG 209  
 QY 190 -----QKVEFKIDIVPRASALPAPPTGSAL 215  
 DB 210 VSKGDRNINVEVERAP-VITVPRRLGQAL 239  
 RESULT 32  
 PIGR RABIT STANDARD; PRT; 773 AA.  
 AC P01832;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polymeric-immunoglobulin receptor precursor (Poly-Ig receptor) (PIGR)  
 DE [Contains: Secretory component].  
 GN PIGR.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OC NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=84142246; PubMed=6322002;  
 RA Mostov K.S., Friedlander M., Blobel G.;  
 RT "The receptor for transendothelial transport of IgA and IgM contains  
 RT multiple immunoglobulin-like domains.";  
 RL Nature 308:37-43(1984).  
 RN [2]  
 RP SEQUENCE OF 87-114 AND 410-428.  
 RX MEDLINE=88228032; PubMed=3131339;  
 RA Fruchtiger S., Hughes G.J., Hanly W.C., Jaton J.-C.;  
 RT "Rabbit secretory components of different allotypes vary in their  
 RT carbohydrate content and their sites of N-linked glycosylation.";  
 RL J. Biol. Chem. 263:8120-8125(1988).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE

CC TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Also secreted.  
 CC -1- POLYMORPHISM: The sequence shown is that of allotype T62.  
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like V-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X00412; CAAG5118.1; -  
 DR PIR; A02111; ORRBG  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003599; Ig.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00409; IG; 5.  
 DR PROSITE; PS50835; IG LIKE; 3.  
 DR Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;  
 KW Polymorphism.  
 FT SIGNAL 1 18  
 FT CHAIN 19 773 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
 FT CHAIN 19 615 SECRETORY COMPONENT.  
 FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 648 670 POTENTIAL.  
 FT DOMAIN 671 773 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 25 131 IG-LIKE V-TYPE 1.  
 FT DOMAIN 138 232 IG-LIKE V-TYPE 2.  
 FT DOMAIN 233 340 IG-LIKE V-TYPE 3.  
 FT DOMAIN 352 455 IG-LIKE V-TYPE 4.  
 FT DOMAIN 461 557 IG-LIKE V-TYPE 5.  
 FT DISULFID 46 115 POTENTIAL.  
 FT DISULFID 155 225 POTENTIAL.  
 FT DISULFID 260 324 POTENTIAL.  
 FT DISULFID 369 438 POTENTIAL.  
 FT DISULFID 478 538 POTENTIAL.  
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (IN ALLOTYPES T61).  
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (IN ALLOTYPES T62  
 FT AND T63).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .).  
 FT VARIANT 88 88 K -> N (IN ALLOTYPES T61).  
 FT VARIANT 94 94 D -> E (IN ALLOTYPES T61).  
 FT VARIANT 101 108 TVDQLTON -> YINRSQS (IN ALLOTYPES T61).  
 FT VARIANT 110 110 S -> T (IN ALLOTYPES T63).  
 SQ SEQUENCE 773 AA; 83886 MW; DF2C44D2F1193C65 CRC64;  
 Query Match 8.5%; Score 112.5; DB 1; Length 773;  
 Best Local Similarity 23.4%; Pred. No. 0.47; 84; Indels 51; Gaps 7;  
 Matches 52; Conservative 35; Mismatches 84; Indels 51; Gaps 7;  
 QY 36 TVELTCTASQKSIQFHWKNSQIKILNQSGFLTKGPKLNDRADSRSLMDQGN 90  
 DB 150 TVTTCFPTVATRLKKS--FYKVEDELVLIIIDSSKKAQD-----RYKGRITLQIQS 202  
 QY 91 ----NPLIINKLKIEDSDTYICE-----VEDQKEVOLVLFGL----- 125  
 DB 203 TTAKFPTVITIKHQLNDAGQVYCCSGSDPRAEQNVRLRLTPELILYGNLGGSYTFECAL 262  
 QY 126 ----TANSDDLHLOQSITLTLSPGSSPSVQCSPPGKNIQGKTLVSQLELODGTW 182  
 DB 263 DSEDANAVASIROGVGNVVIDSGCTDPAFEGILFTKXENHFSVAVIAGLRKEDTGYN 322  
 QY 183 TCTVQLON-----QKVEFKIDIVPRASALPAPPTGS 213  
 DB 323 LCGVQSGSGSGDPTQLRQLFVNEIDVNSPPLKQKFPDGS 364  
 RESULT 33  
 CEAS\_HUMAN STANDARD; PRT; 702 AA.  
 ID CEAS\_HUMAN  
 AC P06731;

```

DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE antigen).
DE CEACAM5 OR CEA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258861; PubMed=2342461;
RA Schreier H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA Haasauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT of its promoter indicates a region conveying cell type-specific
RT expression."
RL Mol. Cell. Biol. 10:2738-2748(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86039876; PubMed=3670312;
RA Beuchtemin N., Benchimol S., Coutroyer D., Fuks A., Stanners C.P.;
RT "Isolation and characterization of full-length functional cDNA clones
RT for human carcinoembryonic antigen."
RL Mol. Cell. Biol. 7:3221-3230(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3220478;
RA Barnett T., Goebel S.J., Northurte M.A., Elting J.J.;
RT "Carcinoembryonic antigen family: characterization of cDNAs coding
RT for NCA and CEA and suggestion of nonrandom sequence variation in
RT their conserved loop-domains."
RL Genomics 3:59-66(1988).
RN [4]
RP SEQUENCE OF 5-702 FROM N.A.
RX MEDLINE=87128144; PubMed=3814146;
RA Oikawa S., Nakazato H., Kosaki G.;
RT "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT from cDNA sequence."
RL Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN [5]
RP SEQUENCE OF 331-702 FROM N.A.
RX MEDLINE=87204247; PubMed=3036371;
RA Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT "Isolation and characterization of cDNA clones encoding the human
RT carcinoembryonic antigen reveal a highly conserved repeating
RT structure."
RL Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
RN [6]
RP SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC derived digestive system epithelium and fetal colon.
CC -1- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 kDa
CC COMPRISING 60% CARBOHYDRATE.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC -1- DATABASE: NAME=FROM; NOTE=CD guide CD66e entry?
CC WWW=http://www.ncbi.nlm.nih.gov/ncbi/cd/cd66e.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17303; AAB59513.1; -.
DR EMBL; M59262; AAB62835.1; ALT. SEQ.
DR EMBL; M59255; AAB62835.1; JOINED.
DR EMBL; M59256; AAB62835.1; JOINED.
DR EMBL; M59257; AAB62835.1; JOINED.

```

[illegible]

Db 544 --SPRLQISNGNRLLTTFNTRNDARAVVCCIONSVSANRSDPYTLVDLYGPDPIIISP 601  
QY 130 DTHLLOGSLTLTLESPGSSPSVOCRRPRKNIQGGKTLVSQLELDGSTMCTVLQN 189  
Db 602 DSSYLSGNLNLSCHSASNPSPOYSWRI-NCIPQHTVLTALITPNNNGTACV--- 657  
QY 190 QKKVEFKIDIVPAPPTGSALPDPQTASALPDPPASALPALVALVSFLGLG 249  
Db 658 -----SNL-ATGRNNSIVKSIIVSASGTSFGLS---AGATVGMIGVALVG 698  
QY 250 VACV 253  
Db 699 VALI 702  
RESULT 34  
ID LAR DROME STANDARD; PRT; 2029 AA.  
AC P16621;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Protein-tyrosine phosphatase Lar precursor (EC 3.1.3.48) (Protein-tyrosine-phosphate phosphohydrolase) (DLAR).  
GN LAR.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RE SEQUENCE FROM N.A.  
RX MEDLINE=90046860; PubMed=2554325;  
RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;  
RT "A family of receptor-linked protein tyrosine phosphatases in humans and Drosophila.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).  
RN [2]  
RE SEQUENCE FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=96178473; PubMed=8598047;  
RA Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S., Saito H.;  
RT "The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in Drosophila.";  
RL Cell 84:611-622(1996).  
CC -1- FUNCTION: It is possible that DLAR is a cell adhesion receptor. It possesses an intrinsic protein tyrosine phosphatase activity (PTPase). It controls motor axon guidance.  
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Selectively expressed in a subset of axons and pioneer neurons in the embryo.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M27700; AAA28668.1; -;  
DR EMBL; U36857; AAC47002.1; -;  
DR EMBL; U36849; AAC47002.1; JOINED.  
DR EMBL; U36850; AAC47002.1; JOINED.  
DR EMBL; U36851; AAC47002.1; JOINED.  
DR EMBL; U36852; AAC47002.1; JOINED.  
DR EMBL; U36853; AAC47002.1; JOINED.

DR EMBL; U36854; AAC47002.1; JOINED.  
DR EMBL; U36855; AAC47002.1; JOINED.  
DR EMBL; U36856; AAC47002.1; JOINED.  
DR PIR; A36182; TDFELK.  
DR HSSP; P28827; IRPM.  
DR FlyBase; FBgn0000464; Lar.  
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IDA.  
DR GO; GO:0008045; P:motor axon guidance; IMP.  
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR00242; TYR\_PP.  
DR Pfam; PF00041; fn3; 9.  
DR Pfam; PF00047; Ig; 3.  
DR Pfam; PF00102; Y\_phosphatase; 2.  
DR PRINTS; PR00014; FNTYPRIII.  
DR PRINTS; PR00700; PRTYPRHTRASE.  
DR SMART; SM00600; FN3; 9.  
DR SMART; SM00408; IGC2; 3.  
DR SMART; SM00194; PTPC; 2.  
DR PROSITE; PS50835; IG\_LIKE; 3.  
DR PROSITE; PS50383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_2.  
KW Hydroxylase; Receptor; Glycoprotein; Signal; Transmembrane; Cell adhesion; Immunoglobulin domain; Repeat.  
FT SIGNAL 1 32  
FT CHAIN 33 2029  
FT DOMAIN 33 1377  
FT TRANSMEM 1378 1402  
FT DOMAIN 1403 2029  
FT FT 36 128  
FT DOMAIN 140 224  
FT FT 234 316  
FT DOMAIN 320 417  
FT DOMAIN 418 512  
FT DOMAIN 513 607  
FT DOMAIN 608 706  
FT DOMAIN 707 809  
FT DOMAIN 810 906  
FT DOMAIN 907 1007  
FT DOMAIN 1008 1102  
FT DOMAIN 1103 1207  
FT DOMAIN 1492 1738  
FT DOMAIN 1781 2029  
FT ACT\_SITE 1670 1670  
FT FT 1961 1961  
FT FT 57 111  
FT DISULFID 161 209  
FT FT 256 301  
FT DISULFID 176 176  
FT CARBOHYD 253 253  
FT CARBOHYD 298 298  
FT CARBOHYD 553 553  
FT CARBOHYD 616 616  
FT CARBOHYD 666 666  
FT CARBOHYD 721 721  
FT CARBOHYD 774 774  
FT CARBOHYD 915 915  
FT CARBOHYD 962 962  
FT CARBOHYD 1183 1183  
FT CARBOHYD 1304 1304  
SO SEQUENCE 2029 AA; 229027 MM; 536AOC794D3DC800 CRC64;  
Query Match 8.4%; Score 110.5; DB 1; Length 2029;  
Best Local Similarity 23.0%; Pred. No. 2.2;  
Matches 50; Conservative 37; Mismatches 75; Indels 55; Gaps 11;

QY 20 PAATGKVVVLGKKDPVELTCTASQKKSIOFMKNSNOIKIGNQGSFLTKGPKSLNDR 79  
 DB 140 PVTGGTGVIEGVHVTMTCKAIGNPTNIYV-INKQTKV-----DM 182  
 QY 80 ADNRSRISMDQGNPLIKNLIKEDSDTYICEVEDEQKEVQLVFGFLTANSDTHL-LOGGS 138  
 DB 183 SNRYSISKD--GF-LOIENREDQGRYECVAENS-----MGTEHSKATNLYKVR 231  
 QY 139 LVTLESPP-----GSSPSVOCRS-----PRGXNIGGKTLT-----VSOL 174  
 DB 232 VPTFSPRPETISEVMIGSNLNLSCIAVGSMPHVKKMKGSEDLTPEENEMPIGRNVQLTI 291  
 QY 175 ELQDSGTCTCTVQGNQKVEFKIDIVPRASALPAPT 211  
 DB 292 NIGESANYTCTIAASTLGQID-SVSVV-KVQSLPTAPT 326

RESULT 35  
 ID VCAL\_MOUSE STANDARD; PRT; 739 AA.  
 AC P29533;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).  
 OS VCAM1 OR VCAM-1.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=FVB; TISSUE=Lung;  
 MD MEDLINE=92181437; PubMed=1371918;  
 RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,  
 RA Burks L., Miyake K., Kincaid P., Lobb R.;  
 RT "Cloning of murine and rat vascular cell adhesion molecule-1.";  
 RL Biochem. Biophys. Res. Commun. 183:163-169(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Lymph node;  
 RX MEDLINE=93246254; PubMed=7683304;  
 RA Araki M., Araki K., Vassalli P.;  
 RT "Cloning and sequencing of mouse VCAM-1 cDNA.";  
 RL Gene 126:261-264(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=129; TISSUE=Embryo;  
 MD MEDLINE=94117008; PubMed=7507076;  
 RA Cybulsky M.I., Allan-Motamed M., Collins T.;  
 RT "Structure of the murine VCAM1 gene.";  
 RL Genomics 18:387-391(1993).  
 RN [4]  
 RP SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).  
 RC STRAIN=NIH Swiss, and 129/Sv;  
 RA Kumar A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.,  
 RA Ballantyne C.M.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB; TISSUE=Lung;  
 MD MEDLINE=93232042; PubMed=7682556;  
 RA Moy P., Lobb R., Tizard R., Olson D., Hession C.;  
 RT "Cloning of an inflammation-specific phosphatidyl inositol-linked  
 RT form of murine vascular cell adhesion molecule-1.";  
 RL J. Biol. Chem. 268:8835-8841(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=95015899; PubMed=7523515;  
 RA Kumar A.G., Dai X.Y., Kozak C.A., Mims M.P., Gotto A.M.,  
 RA Ballantyne C.M.;

RT "Murine VCAM-1. Molecular cloning, mapping, and analysis of a  
 RT truncated form.";  
 RT J. Immunol. 153:4088-4096(1994).  
 RN [7]  
 RP SEQUENCE OF 311-345 FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=93317595; PubMed=7687058;  
 RA Terry R.W., Kwee L., Levine J.F., Labow M.A.;  
 RT "Cytokine induction of an alternatively spliced murine vascular cell  
 RT adhesion molecule (VCAM) mRNA encoding a  
 RT glycosylphosphatidylinositol-anchored VCAM protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5919-5923(1993).  
 RN [8]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC TISSUE=Endothelial cells;  
 RA Korenaga R., Ando J., Tsuboi H., Kamiya A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
 CC IN LEUCOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
 CC INTEGRIN VLA4 ON LEUCOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL  
 CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A  
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUCOCYTE  
 CC EMIGRATION TO SITES OF INFLAMMATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG ISOFORM) OR  
 CC ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT ISOFORM).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=Long;  
 CC IsoId=P29533-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Short;  
 CC IsoId=P29533-2; Sequence=VSP\_002581, VSP\_002582;  
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
 CC well as on macrophage-like and dendritic cell types in both normal  
 CC and inflamed tissue.  
 CC -1- PTM: The GPI-anchor is located on position 319 of isoform 2.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domain.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL; W64487; AAA40545.1; -  
 CC EMBL; X67783; CAA47989.1; -  
 CC EMBL; L22355; AAA16921.1; JOINED.  
 CC EMBL; L22301; AAA16921.1; JOINED.  
 CC EMBL; L22349; AAA16921.1; JOINED.  
 CC EMBL; L22350; AAA16921.1; JOINED.  
 CC EMBL; L22351; AAA16921.1; JOINED.  
 CC EMBL; L22352; AAA16921.1; JOINED.  
 CC EMBL; L22353; AAA16921.1; JOINED.  
 CC EMBL; L22354; AAA16921.1; JOINED.  
 CC EMBL; L22350; AAA16920.1; -  
 CC EMBL; L22301; AAA16920.1; JOINED.  
 CC EMBL; L22349; AAA16920.1; JOINED.  
 CC EMBL; U12878; AAB60659.1; ALT\_SEQ.  
 CC EMBL; U12879; AAB60660.1; ALT\_SEQ.  
 CC EMBL; U12880; AAB60661.1; ALT\_SEQ.  
 CC EMBL; U12874; AAB60662.1; ALT\_SEQ.  
 CC EMBL; U12875; AAB60663.1; ALT\_SEQ.  
 CC EMBL; U12883; AAB60664.1; ALT\_SEQ.  
 CC EMBL; U12881; AAA80010.1; ALT\_SEQ.  
 CC EMBL; U12882; AAA80011.1; ALT\_SEQ.  
 CC EMBL; U12875; AAA80012.1; ALT\_SEQ.  
 CC EMBL; U12872; AAA80013.1; ALT\_SEQ.  
 CC EMBL; U12876; AAA80014.1; ALT\_SEQ.  
 CC EMBL; U12873; AAA80015.1; ALT\_SEQ.  
 CC EMBL; U12877; AAA80016.1; ALT\_SEQ.  
 CC EMBL; U108431; AAA40546.1; -  
 CC EMBL; U12884; AAA64832.1; -

```

DR EMBL; L12541; AAC37607.1; -
DR EMBL; U42327; AAB8576.1; -
DR PIR; B46819; A46052.
DR PIR; JN0581; JN0581.
DR HSSP; P19320; IVCA.
DR MGI; MGI:98926; Vcam1.
DR GO; GO:0007155; P:cell adhesion, IDA.
DR InterPro; IPR003987; ICM Vcam-1.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003598; Ig-C2.
DR InterPro; IPR003989; Vcam-1.
DR Pfam; PF00047; Ig_5.
DR PRINTS; PRO1472; ICMVcam1.
DR PRINTS; PRO1474; Vcam1.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS50835; IG_LIKE; 5.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
Repeat; GPI-anchor; Signal; Alternative splicing.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.
FT TRANSMEM 699 720 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 119 212 IG-LIKE C2-TYPE 1.
FT DOMAIN 223 309 IG-LIKE C2-TYPE 2.
FT DOMAIN 312 393 IG-LIKE C2-TYPE 3.
FT DOMAIN 408 506 IG-LIKE C2-TYPE 4.
FT DOMAIN 511 595 IG-LIKE C2-TYPE 5.
FT DOMAIN 600 682 IG-LIKE C2-TYPE 6.
FT DISULFID 47 95 BY SIMILARITY.
FT DISULFID 52 95 BY SIMILARITY.
FT DISULFID 137 195 BY SIMILARITY.
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASAPLIC 310 345 EKFPTVDSFGSQAQVGVDSVLTCAIGCDSPSF -> D
GRMSQITNGHQLTVHLMFAKSFYCYLCVLYAL (1n
isoform 2).
FT VASAPLIC 346 739 /FtId=VSP 002581.
FT CONFLICT 693 693 Missing (In isoform 2).
FT SEQUENCE 739 AA; 81317 MW; 302134C341E5B449 CRC64;
Query Match 8.4%; Score 110; DB 1; Length 739;
Best Local Similarity 21.3%; Pred. No. 0.69;
Matches 40; Conservative 34; Mismatches 68; Indels 46; Gaps 6;
QY 32 KKGDVELTCTASQKSIQFMW--KNSNOI-KILGNQGSFLTQKPSKLNDRADSRRLMD 88
DB 237 QEGGAVMTCTSGEGIPAEIFMGRKLDNEVLQL----- 270
QY 89 QGNFPLIKNIKIEDSDTYICE---VEDQEEVQLLV-----FGLTANSDTLLOQGS 138
DB 271 SGNATLTLILAMRMBDSGYVEGVNLIGRDABEVLVQEKPTVIDSPSGQVAAYGDS 330
QY 139 LTLTLESPGSSPVSQCRSPRKNIOG-----GKTLVSQLELDLQSGTWTCTVLQNOK 191
DB 331 VVLTCALIGCDSPFSFMRKTQTDSPLVNGVNRGAKSTLVLSVGFEDSHSLCAVTCLQR 390
QY 192 KVEFKIDI 199
DB 391 TLEKRTQV 398

```

```

RESULT 36
AMAL DROME STANDARD; PRT; 333 AA.
AC P15364; QGV3AS;
DT 01-APR-1990 (Rel. 14, Created)

```

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amalgam protein precursor
GN AMA OR BG:DS00276.6 OR CG2198.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pelecyota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=89028670; PubMed=3141062;
RA Seeger M.A., Haffley L., Kaufman T.C.;
RT "Characterization of amalgam: a member of the immunoglobulin
RL superfamily from Drosophila."
RL Cell 55:589-600(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Celniker S.E., Pfeiffer B.D., Kafets J., Martin C.H., Mayeda C.A.,
RT Palazzolo M.J.;
RT "Complete sequence of the Antennapedia complex of Drosophila."
RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hughes R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spryer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommler B., Paclob J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;

```



RT (CAM-L1).";  
 RL Nucleic Acids Res. 19:5395-5401(1991).  
 RN [10]  
 RP SEQUENCE OF 809-1257 FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Fetsch E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marzella K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stajich M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uadit T.B., Toshiyuki S., Carinici P., Pearce C.,  
 RA Bata S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP SEQUENCE OF 1030-1257 FROM N.A.  
 RX MEDLINE=9113183; PubMed=1993895;  
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,  
 RA Stalcup W.B.;  
 RT "Isolation and sequence of partial cDNA clones of human LI: homology  
 of human and rodent LI in the cytoplasmic region.";  
 RL J. Neurochem. 56:797-804(1991).  
 RN [12]  
 RP VARIANT HSAS TYR-264.  
 RX MEDLINE=94004956; PubMed=8401576;  
 RA Jouet M., Rosenthal A., Macfarlane J., Kenrick S., Donnai D.;  
 RT "A missense mutation confirms the LI defect in X-linked hydrocephalus  
 (HSAS).";  
 RL Nat. Genet. 4:331-331(1993).  
 RN [13]  
 RP VARIANT HSAS/MASA LEU-1194.  
 RX MEDLINE=95187172; PubMed=7881431;  
 RA Fransen E., Schander-Stumpel C., Vits L., Coucke P., van Camp G.,  
 RA Willems P.J.;  
 RT "X-linked hydrocephalus and MASA syndrome present in one family are  
 due to a single missense mutation in exon 28 of the LICAM gene.";  
 RL Hum. Mol. Genet. 3:2255-2256(1994).  
 RN [14]  
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.  
 RX MEDLINE=95004608; PubMed=7920659;  
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,  
 RA Paterson J., Metzberg A., Ionescu V., Temple K., Kenrick S.;  
 RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked  
 hydrocephalus result from mutations in the LI gene.";  
 RL Nat. Genet. 7:402-407(1994).  
 RN [15]  
 RP VARIANTS MASA GLN-210 AND ASN-598.  
 RX MEDLINE=95004609; PubMed=7920660;  
 RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,  
 RA Reijnders E., Korn B., Pousetka A., Wilson G., Schander-Stumpel C.,  
 RA Winter R.M., Schwartz C., Willems P.J.;  
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene  
 LICAM.";  
 RL Nat. Genet. 7:408-413(1994).  
 RN [16]  
 RP VARIANTS HSAS/MASA SER-9; SER-121; LYS-309; PHE-768; LEU-941 AND  
 RX MEDLINE=95283776; PubMed=7762552;  
 RA Jouet M., Moncla A., Paterson J., McKewen C., Fryer A., Carpenter N.,  
 RA Holmberg E., Wadellius C., Kenrick S.;  
 RT "New domains of neural cell-adhesion molecule LI implicated in

RT X-linked hydrocephalus and MASA syndrome.";  
 RL Am. J. Hum. Genet. 56:1304-1314(1995).  
 RN [17]  
 RP VARIANTS HSAS/MASA GLN-184; GLN-210; TYR-264; ARG-452; ASN-598 AND  
 RX MEDLINE=96151146; PubMed=8556302;  
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;  
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,  
 RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus  
 RT due to mutations in one single gene, LI.";  
 RL Eur. J. Hum. Genet. 3:273-284(1995).  
 RN [18]  
 RP ERRATUM.  
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;  
 RL Eur. J. Hum. Genet. 4:126-126(1996).  
 RN [19]  
 RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.  
 RX MEDLINE=96057511; PubMed=7562969;  
 RA Ruiz J.C., Cuppens H., Legius E., Fyngs J.-P., Glover T., Marynen P.,  
 RA Casman J.-J.;  
 RT "Mutations in LI-CAM in two families with X linked complicated  
 RT spastic paraplegia, MASA syndrome, and HSAS.";  
 RL J. Med. Genet. 32:549-552(1995).  
 RN [20]  
 RP VARIANTS HSAS CYS-194 AND LEU-240.  
 RX MEDLINE=97083370; PubMed=8929944;  
 RA Gu S.-M., Orich U., Veske A., Enders H., Kluender K., Schloesser M.,  
 RA Engel W., Schwinger E., Gal A.;  
 RT "Five novel mutations in the LICAM gene in families with X linked  
 RT hydrocephalus.";  
 RL J. Med. Genet. 33:103-106(1996).  
 RN [21]  
 RP VARIANTS HSAS GLN-184; 439-VAL--THR-443 DEL; CYS-784 AND  
 RX 936-LEU-LEU-948 DEL.  
 RX MEDLINE=97338664; PubMed=9195224;  
 RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,  
 RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goodie D.,  
 RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenrick S.;  
 RT "Nine novel LI CAM mutations in families with X-linked  
 RT hydrocephalus.";  
 RL Hum. Mutat. 9:512-518(1997).  
 RN [22]  
 RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.  
 RX MEDLINE=98180721; PubMed=9521424;  
 RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;  
 RT "Multiple exon screening using restriction endonuclease  
 RT fingerprinting (REF) detection of six novel mutations in the LI cell  
 RT adhesion molecule (LICAM) gene.";  
 Query Match 8.3%; Score 109; DB 1; Length 1257;  
 Best Local Similarity 20.9%; Pred. No. 1.6;  
 Matches 46; Conservative 35; Mismatches 97; Indels 50; Gaps 9;  
 Oy 8 RHLLV-----LQALPAA--ATGCKVVLKKGDTVLTCTASOKSIOFMKNSNQIKI 61  
 Db 409 RHGLLLANAYIVYVQPLAKLITLADNQTVMVQGSTAYLLKAGAPVPSQMLDEDTTV 468  
 Oy 62 LGNGSEFLTKGPKLNDRADRSRLMDQGNFPLIKNLKIEDSPYICEVEDQKEVQL 121  
 Db 469 LGDERF-----PYANGTIGIRLQANDGRYCCLANQNNVTIM 509  
 Oy 122 VFGITANSPTHLQ-----GQSITLLES--PGSSPSVQCKSPRGKNIQ----- 164  
 Db 510 A-MLKVDAAQIOTGPGSTIEKKGRVTFPCASFPSPSLPSTWRD-CRDIQELQSD 567  
 Oy 165 -----GGKTSVSGLELDSDGTCTQVQKQKVEFIDIVPAPASALPAP 209  
 Db 568 KYFLIEDGR-LVHSLDSDGNTSCVASTELDVESHQAQLLVGSPQVP 616  
 RESULT 38  
 VGR2 MOUSE  
 ID\_VGR2\_MOUSE STANDARD; PRT; 1367 AA.

AC p35918;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular endothelial growth factor receptor 2 precursor (EC 2.7.1.112)  
 DE (VEGFR-2) (Protein-tyrosine kinase receptor flk-1) (Fetal liver kinase  
 1) (kinase NKX).  
 GN KDR OR FLK1 OR FLK-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Embryo;  
 RX MEDLINE=93208880; Pubmed=7681362;  
 RA Millaer B., Witzmann-Voos S., Schnurch H., Martinez R.,  
 RA Weiler N.P.H., Risau W., Ullrich A.;  
 RT "High affinity VEGF binding and developmental expression suggest  
 RT Flk-1 as a major regulator of vasculogenesis and angiogenesis.";  
 RL Cell 72:835-846(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/He; TISSUE=fetal liver;  
 RX MEDLINE=92020984; Pubmed=1717995;  
 RA Mathews W., Jordan C.T., Gavin M., Jenkins N.A., Copeland N.G.,  
 RA Lemischka I.R.;  
 RT "A receptor tyrosine kinase cDNA isolated from a population of  
 RT enriched primitive hematopoietic cells and exhibiting close genetic  
 RT linkage to c-kit.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9026-9030(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93141255; Pubmed=8423988;  
 RA Oelrichs R.B., Reid H.H., Bernard O., Ziemiecki A., Wilks A.F.;  
 RT RT receptor for vascular endothelial growth factor isolated from  
 RT E10 embryonic neuroepithelium is expressed in endothelial cells of  
 RT the developing embryo.";  
 RL Oncogene 8:11-18(1993).  
 RN [4]  
 RP SEQUENCE OF 1-15 FROM N.A.  
 RX MEDLINE=96032749; Pubmed=7559454;  
 RA Patterson C., Petrella M.A., Hsieh C.-M., Yoshizumi M., Lee M.-E.,  
 RA Harber E.;  
 RT "Cloning and functional analysis of the promoter for KDR/flk-1, a  
 RT receptor for vascular endothelial growth factor.";  
 RL J. Biol. Chem. 270:23111-23118(1995).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE=93361481; Pubmed=8356051;  
 RA Quinn T.P., Peters K.G., de Vries C., Ferrara N., Williams L.T.;  
 RT "Fetal liver kinase 1 is a receptor for vascular endothelial growth  
 RT factor and is selectively expressed in vascular endothelium.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7533-7537(1993).  
 CC -1- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN  
 CC KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM  
 CC PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF  
 CC VASCULAR PERMEABILITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN ADULT HEART, LUNG,  
 CC KIDNEY, BRAIN AND SKELETAL MUSCLE, BUT IS ALSO EXPRESSED AT LOWER  
 CC LEVELS IN MOST OTHER ADULT TISSUES.  
 CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.  
 CC CSF-1/PDGF receptor subfamily.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 209  
 CC FT DOMAIN 226 325  
 CC FT DOMAIN 330 416  
 CC FT DOMAIN 423 542  
 CC FT DOMAIN 549 656  
 CC FT DOMAIN 665 751  
 CC FT DOMAIN 832 1160  
 CC FT NP\_BIND 838 846  
 CC FT BINDING 866 866  
 CC FT ACT\_SITE 1026 1026  
 CC FT CARBOHYD 46 46  
 CC FT CARBOHYD 98 98  
 CC FT CARBOHYD 145 145  
 CC FT CARBOHYD 160 160  
 CC FT CARBOHYD 247 247  
 CC FT CARBOHYD 320 320  
 CC FT CARBOHYD 376 376  
 CC FT CARBOHYD 397 397  
 CC FT CARBOHYD 509 509  
 CC FT CARBOHYD 521 521  
 CC FT CARBOHYD 578 578  
 CC FT CARBOHYD 611 611  
 CC FT CARBOHYD 617 617  
 CC FT CARBOHYD 629 629  
 CC FT CARBOHYD 673 673  
 CC FT CARBOHYD 702 702  
 CC FT CARBOHYD 719 719  
 CC FT MOD\_RES 1057 1057  
 CC FT CONFLICT 25 25  
 CC FT CONFLICT 679 679  
 CC FT CONFLICT 783 783  
 CC FT CONFLICT 917 917  
 CC FT CONFLICT 1341 1367  
 CC FT 1367 AA; 152516 MW; EFC99704F1DCA266 CRC64;  
 CC SEQUENCE  
 CC -----  
 CC entires requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X70842; CAA50192.1; -  
 CC DR EMBL; X59397; CAA42040.1; -  
 CC DR EMBL; S53103; AAB35043.1; -  
 CC DR EMBL; X89777; CAA61917.1; -  
 CC DR PIR; A41228; A41228.  
 CC HSSP; P11362; IFGK.  
 CC MGD; MGI:96683; Kdr.  
 CC GO; GO:0045165; Pcell fate commitment; IMP.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR001824; ReceptTykKinIII.  
 CC InterPro; IPR001245; Tyr\_kinase.  
 CC InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC Pfam; PF00047; Ig\_6.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 2.  
 CC SMART; SM00408; Igc2; 1.  
 CC SMART; SM00219; TyKc; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 CC KW Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;  
 CC KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 CC Immunoglobulin domain; Repeat.  
 CC FT CHAIN 1 19  
 CC FT 1 1367  
 CC FT 20 762  
 CC FT TRANSMEM 763 784  
 CC FT DOMAIN 785 1367  
 CC FT DOMAIN 146 111  
 CC FT DOMAIN 143 2



```

Query Match Similarity      8.3%; Score 109; DB 1; Length 1367;
Best Local Similarity     21.0%; Pred. No. 1.8;
Matches    45; Conservative   36; Mismatches    73; Indels    60; Gaps    9

Oy          HLLVLQALLPAAATGKNNVVGKKGDPTVELCTASOKKSIGPHKNNSNQIKLNGQS F 68
Db          HVINGELTVGPAPAQPTQ-----EVSILCTIDRNTFEULTRYK-----LSQATS 590
Oy          69 LTKGPSKLNDRAADSRSIM-----DGNFPLII--KNLKIEDSDTYICEVEDQKEB- 117
Db          591 VHMGES-LTPVCNKINDALWKLTNGTFWSNSTNIDLIVAFQNASTLDQGDDVCAQDKYTKX 649
Oy          118 -----VOLLVGLTNSDTHLGGSLTLTESPPGSBSVQCSPRKCN-----         162
Db          650 RHCLVKQLILERMAPMTIGNLENSTTTI-----GETIEVTCC-PAGNPPTHITWPD 701
Oy          163 -----IQSGKTLSVSQLDPSGTWC 184
Db          702 NETIVEDSGYLVRDGNRLTIIRRVAKEDGELYTC 735

RESULT 39
PCBM_HUMAN        STANDARD:       PRT:   4391 AA.
AC              P98160.Q16287.Q9H3V5;
DT              01-OCT-1996 (Rel. 34, Created)
DT              28-FEB-2003 (Rel. 41, Last sequence update)
DT              10-OCT-2003 (Rel. 42, Last annotation update)
DE              Basement membrane-specific heparan sulfate proteoglycan core
DE              protein precursor (HSPG) (perlecan) (PLOC).
OS              Homo sapiens (Human).
OC              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
OC              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX              NCBI_TaxId=9606;
[1]
RP              SEQUENCE FROM N.A. MEDLINE=92112994; PubMed=1730768;
RX              Kallunki P., Tryggvason K.;
RT              "Human basement membrane heparan sulfate proteoglycan core protein: a
RT              467-kD protein containing multiple domains resembling elements of the
RT              low density lipoprotein receptor, laminin, neural cell adhesion
RT              molecules, and epidermal growth factor.";
RL              J. Cell Biol. 116:559-571(1992).
[2]
RN              SEQUENCE FROM N.A. TISSUE=Codon, And Skin;
RC              MEDLINE=92235084; PubMed=1569102;
RX              Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT              "Primary structure of the human heparan sulfate proteoglycan from
RT              basement membrane (HSPG/perlecan). A chimeric molecule with multiple
RT              domain homologous to the low density lipoprotein receptor, laminin,
RT              neural cell adhesion molecules, and epidermal growth factor.";
RL              J. Biol. Chem. 267:8544-8557(1992).
[3]
RN              SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJSI TYR-1532.
RX              MEDLINE=20553141; PubMed=1101850;
RX              Nicole S., Davoine C.-S., Topaloglu H., Cartoglio L., Barral D.,
RX              Belgheon P., Ben-Hamida C., Hamouda H., Craud C., White P.S.,
RX              Sanson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RX              Henlert F., Fontaine B.;
RT              "Perlecan, the major proteoglycan of basement membranes, is altered in
RT              patients with Schwartz-Jampel syndrome (chondrodysplastic myotonia).";
NL              Nat. Genet. 26:480-483(2000).
[4]
RP              SEQUENCE OF 1016-1470 FROM N.A. TISSUE=Codon;
RX              MEDLINE=91365376; PubMed=1679749;
RX              Dodge G.F., Kovalevsky I., Chu M.L., Hassell J.R., McBride O.W.,
RA              Yi H.G., Iozzo R.V.;
RT              "Heparan sulfate proteoglycan of human colon: partial molecular
RT              cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT              short arm of human chromosome 1.";
```

```

RL Genomics 10:673-680(1991).
RN [5]
RN SEQUENCE OF 890-1396 FROM N.A.
RX TISSUE=Fibrosarcoma;
RC MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Trygvasson K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1->p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [6]
RN SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessell S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RN CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membranes.
CC -1- PM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC characterized by permanent myotonia (prolonged failure of muscle
CC relaxation) and skeletal dysplasia, resulting in reduced stature,
CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC -1- SIMILARITY: Contains 4 LBD-receptor class A domains.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin IV domains.
CC -1- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X62515; CAA44373.1; -
DR EMBL, M85289; AAA52700.1; -
DR EMBL, AL445795; CAC18534.1; -
DR EMBL, M64283; AAA52699.1; -
DR EMBL, S76436; AAB21121.2; -
DR EMBL, L22078; -; NOT_ANNOTATED_CDS.
DR PIR, A38096; A38096.
DR HSSP, P00740; 1EDM.
DR Slens-2DPAGE; P98160; -
DR Genew; HGNC:5273; HSPG2.
DR MIM; 142461; -
DR MIM; 255800; -
DR InterPro; IPR008985; Cona_like_1ec_g1.
DR InterPro; IPR007742; EGF_2.
DR InterPro; IPR006209; EGF-like.

```

DR InterPro: IPR006210; IEGF.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR InterPro: IPR003596; Ig\_v.  
 DR InterPro: IPR000034; Laminin\_B.  
 DR InterPro: IPR002048; Laminin\_EGF.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR002172; LDL\_receptor\_A.  
 DR InterPro: IPR000082; SEA\_domain.  
 DR Pfam: PF00008; EGF\_4.  
 DR Pfam: PF00047; Ig\_22.  
 DR Pfam: PF00052; Laminin\_B\_3.  
 DR Pfam: PF00053; Laminin\_EGF\_7.  
 DR Pfam: PF00054; Laminin\_G\_3.  
 DR Pfam: PF00057; Idl\_recept\_a; 4.  
 DR Pfam: PF01390; SEA\_1.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRODOM: PD003031; Laminin\_B\_3.  
 DR SMART: SM00181; EGF\_15.  
 DR SMART: SM00180; EGF\_Lam; 12.  
 DR SMART: SM00409; IG\_22.  
 DR SMART: SM00408; IGC2; 21.  
 DR SMART: SM00406; IGV\_7.  
 DR SMART: SM00281; Lamb; 3.  
 DR SMART: SM00282; LambG; 3.  
 DR SMART: SM00192; LDLa; 4.  
 DR PROSITE: PS00022; EGF\_1; 9.  
 DR PROSITE: PS01186; EGF\_2; 6.  
 DR PROSITE: PS50026; EGF\_3; 4.  
 DR PROSITE: PS50835; IG\_LIKE; 22.  
 DR PROSITE: PS50025; LAM\_G\_DOMAIN; 3.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 11.  
 DR PROSITE: PS01209; LDLRA\_1; 4.  
 DR PROSITE: PS50068; LDLRA\_2; 4.  
 DR PROSITE: PS50024; SEA\_1.  
 KM Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;  
 KM Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;  
 KM Extracellular matrix; EGF-like domain; Disease mutation.  
 FT SIGNAL 1 21  
 FT CHAIN 22 4391  
 FT DOMAIN 80 194  
 FT DOMAIN 198 235  
 FT DOMAIN 284 320  
 FT DOMAIN 324 360  
 FT DOMAIN 367 404  
 FT DOMAIN 405 504  
 FT DOMAIN 521 530  
 FT DOMAIN 531 730  
 FT DOMAIN 731 763  
 FT DOMAIN 764 813  
 FT DOMAIN 814 871  
 FT DOMAIN 879 923  
 FT DOMAIN 924 933  
 FT DOMAIN 934 1125  
 FT DOMAIN 1126 1158  
 FT DOMAIN 1159 1208  
 FT DOMAIN 1209 1265  
 FT DOMAIN 1275 1324  
 FT DOMAIN 1325 1334  
 FT DOMAIN 1335 1329  
 FT DOMAIN 1330 1562  
 FT DOMAIN 1530 1612  
 FT DOMAIN 1613 1670  
 FT DOMAIN 1671 1771  
 FT DOMAIN 1772 1865  
 FT DOMAIN 1866 1955  
 FT DOMAIN 1956 2051  
 FT DOMAIN 2052 2151  
 FT DOMAIN 2152 2244  
 FT DOMAIN 2245 2340

FT DOMAIN 2341 2436 IG-LIKE C2-TYPE 9.  
 FT DOMAIN 2437 2533 IG-LIKE C2-TYPE 10.  
 FT DOMAIN 2534 2629 IG-LIKE C2-TYPE 11.  
 FT DOMAIN 2630 2726 IG-LIKE C2-TYPE 12.  
 Query Match 8.3%; Score 109; DB 1; Length 4391;  
 Best local similarity 19.5%; Pred. No. 7.4; Indels 116; Gaps 11;  
 Matches 63; Conservative 36; Mismatches 108; Indels 116; Gaps 11;  
 QY 12 LVLTALLPAAATQGNKVLGGKGDVLTCTASQKKS1QFHWKN-----SNQI- 59  
 DB 3291 IILHVSPPYATVTPHSAVSQGETVQLQCLAHGFPPLTFQMSRVGSLPGRATANNEL 3350  
 QY 60 -----KILNQSS-----FLTKGP----- 73  
 DB 3351 HEPAPEDSGRYRCVTKKVSAAEFAQLLVQGPFGSLPATSIIPAGSPPTVQVTOLET 3410  
 QY 74 -----SKLNDRAISRSLMDQGNF-----LIIKLIKIDSPYICEVE- 112  
 DB 3411 KSIIGASVEFHCAVPSDRGTQLMFKEGGQLPFGHSGQDGLVLAIQNLDSGCGGTICQAHG 3470  
 QY 113 -----DQKEVQLVFGLTANSPTLLIQGSLTTLTSPGSSP 151  
 DB 3471 PMGKAQASALVLYQALPSVLINIRTSVQTVVVG-----HAVEFCLALGDPKPGVWTS 3523  
 QY 152 SVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVQ-----NKKVEFKIDIVRASALP 207  
 DB 3524 KVGGLRPG-IVQSGGVVRIHVELADAGVRCATNAAGTTQSHVLLVQALPOLIS-MP 3581  
 QY 208 AP---PTGSALPDPQTASALPDP 227  
 DB 3582 QEVVPAGSAVAFPCIASGYTP 3604  
 RESULT 40  
 FCEA\_HUMAN STANDARD; PRT; 257 AA.  
 ID FCEA\_HUMAN  
 AC P12319;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE High affinity immunoglobulin epsilon receptor alpha-subunit precursor  
 DE (PCEB) (IgE Fc receptor, alpha-subunit) (Fc-epsilon RI-alpha).  
 GN PCEB1A OR FCE1A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88233953; PubMed=2967464;  
 RA Kochan J., Pettine L.F., Hakimi J., Kishi K., Kinet J.-P.;  
 RT "Isolation of the gene coding for the alpha subunit of the human high  
 affinity IgE receptor.";  
 RL Nucleic Acids Res. 16:3584-3584(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Mast cells;  
 RX MEDLINE=88158102; PubMed=2964640;  
 RA Shimizu A., Tepler I., Benfey P.N., Berenstein E.H., Siraganian R.P.,  
 RA Leder P.;  
 RT "Human and rat mast cell high-affinity immunoglobulin E receptors:  
 RT characterization of putative alpha-chain gene products.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1907-1911(1988).  
 RN [3]  
 RP 3D-STRUCTURE MODELING OF 26-197.  
 RX MEDLINE=93113350; PubMed=1472946;  
 RA Padlan E.A., Helm B.A.;  
 RT "A modeling study of the alpha-subunit of human high-affinity  
 RT receptor for immunoglobulin-E.";  
 RL Receptor 2:129-144(1992).  
 CC -I- FUNCTION: BINDS TO THE FC REGION OF IMMUNOGLOBULIN EPSILON. HIGH  
 AFFINITY RECEPTOR. RESPONSIBLE FOR INITIATING THE ALLERGIC

CC RESPONSE. BINDING OF ALLERGEN TO RECEPTOR-BOUND IGE LEADS TO CELL  
 CC ACTIVATION AND THE RELEASE OF MEDIATORS (SUCH AS HISTAMINE)  
 CC RESPONSIBLE FOR THE MANIFESTATIONS OF ALLERGY. THE SAME RECEPTOR  
 CC ALSO INDICES THE SECRETION OF IMPORTANT LYMPHOKINES.  
 CC -1- SUBUNIT: TETRAMER OF AN ALPHA CHAIN, A BETA CHAIN, AND TWO  
 CC DISULFIDE LINKED GAMMA CHAINS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 DR EMBL; X06948; CA30025.1; -;  
 DR EMBL; J03605; AAA36204.1; -;  
 DR EMBL; A21606; CAA01564.1; -;  
 DR PIR; S00682; S00682;  
 DR PDB; 1ALS; 27-FEB-95.  
 DR PDB; 1ALT; 27-FEB-95.  
 DR PDB; 1F2Q; 08-JUN-00.  
 DR PDB; 1J86; 29-AUG-01.  
 DR PDB; 1J87; 29-AUG-01.  
 DR PDB; 1J89; 29-AUG-01.  
 DR PDB; 1J89; 05-SEP-01.  
 DR Genew; HGNC:3609; FCER1A.  
 DR MIM; 147140; -;  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR Pfam; PF00047; Ig; 2.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 2.  
 DR IGF-binding protein; Receptor; Transmembrane; Glycoprotein; Signal;  
 KM Immunoglobulin domain; Repeat; 3d-structure.  
 FT SIGNAL 1  
 FT CHAIN 25  
 FT 26 257 HIGH AFFINITY IMMUNOGLOBULIN EPSILON  
 FT DOMAIN 26 205 RECEPTOR ALPHA-SUBUNIT.  
 FT TRANSMEM 226 224 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 225 257 POTENTIAL.  
 FT DOMAIN 30 110 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 111 193 IG-LIKE 1.  
 FT DISULFID 51 93 IG-LIKE 2.  
 FT CARBOHYD 46 67 BY SIMILARITY.  
 FT CARBOHYD 46 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 67 75 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 31 35  
 FT HELIX 39 42  
 FT TURN 43 43  
 FT STRAND 48 54  
 FT STRAND 61 66  
 FT TURN 67 68  
 FT STRAND 69 71  
 FT STRAND 74 74  
 FT TURN 75 76  
 FT STRAND 77 79  
 FT HELIX 82 87  
 FT STRAND 91 96  
 FT TURN 98 100  
 FT HELIX 101 101  
 FT STRAND 104 107  
 FT STRAND 114 116  
 FT HELIX 120 124  
 FT STRAND 128 133

FT STRAND 143 145  
 FT HELIX 146 148  
 FT TURN 149 149  
 FT STRAND 150 150  
 FT TURN 152 153  
 FT STRAND 154 155  
 FT TURN 156 159  
 FT STRAND 160 163  
 FT HELIX 165 170  
 FT STRAND 175 178  
 FT STRAND 181 182  
 FT TURN 183 185  
 FT STRAND 186 187  
 FT STRAND 190 193  
 SQ SEQUENCE 257 AA; 29596 MW; F183BB2357DDAD58 CRC64;

Query Match 8.28; Score 108; DB 1; Length 257;  
 Best Local Similarity 27.18; Pred. No. 0.27;  
 Matches 55; Conservative 22; Mismatches 64; Indels 62; Gaps 10;

QY 10 LLLVLLQLALP-----AAQGNKVYLQK-----KGDVVELTCTASQKKSIQFHKNKSNQI 59  
 DB 10 LLLVLLFPADGVLAVPQKVSINPNNRIFKGVNTLTC----- 51  
 QY 60 KILGNQSFLLTKGPKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVE--DQKEE 117  
 DB 52 ----NGNNFPEVSTKPFHNG-----SLSEETNSLNIVNAKFEDSGEYKCOHQVNESEP 103  
 QY 118 VOLLVPGIYLTNSDTHLLQGSLLTTLTSPGSSPSVQCRSPRGKNI-----QGGKTL-- 169  
 DB 104 VYLEVF-----SDLLLLQA-SAEVMEGQP---LFLRCHGRMNDVYKVIYKDGEMALKY 154  
 QY 170 -----SVSOLHELDSDGTWTCT 185  
 DB 155 WYENHNISTNATVEDSGTYTCT 177

RESULT 41  
 A33\_HUMAN STANDARD; PRT; 319 AA.  
 ID A33\_HUMAN  
 AC Q99795;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Cell surface A33 antigen precursor (Glycoprotein A33).  
 GN GPA33.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Colon carcinoma;  
 RX MEDLINE=97165045; PubMed=9012807;  
 RA Heath J.K., White S.J., Johnstone C.N., Catimel B., Simpson R.J.,  
 RA Moritz R.L., Tu G.-F., Ji H., Whitehead R.H., Greene L.C.,  
 RA Scott A.W., Ritter G., Cohen L., White S., Old L.J., Nice E.C.,  
 RA Burgess A.W.;  
 RA "The human A33 antigen is a transmembrane glycoprotein and a novel  
 RT member of the immunoglobulin superfamily.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:469-474(1997).  
 RN [2]  
 RP POST-TRANSLATIONAL MODIFICATIONS.  
 RX MEDLINE=9736159; PubMed=9245713;  
 RA Ritter G., Cohen L.S., Nice E.C., Catimel B., Burgess A.W.,  
 RA Moritz R.L., Ji H., Heath J.K., White S.J., Welt S., Old L.J.,  
 RA Simpson R.J.;  
 RT "Characterization of posttranslational modifications of human A33  
 RT antigen, a novel palmitoylated surface glycoprotein of human  
 RT gastrointestinal epithelium.";  
 RL Biochem. Biophys. Res. Commun. 236:682-686(1997).  
 CC -1- FUNCTION: May play a role in cell-cell recognition and signaling.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -1- TISSUE SPECIFICITY: Expressed in normal gastrointestinal
CC epithelium and in 95% of colon cancers.
CC -1- PTM: N-GLYCOSYLATED, CONTAINS APPROXIMATELY 8 KDA OF N-LINKED
CC CARBOHYDRATE.
CC -1- PTM: Palmitoylated.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U79725; AAC50957.1; -.
CC Genew; HGNC:4445; GPA33.
CC MIM; 602171; -.
CC DR GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
CC DR GO; GO:0004872; F:receptor activity; TAS.
CC DR InterPro; IPR007110; IG_V.
CC DR Pfam; PF00047; IG_2.
CC DR SMART; SM00406; IG_1.
CC DR PROSITE; PSS0835; IG_LIKE; 2.
CC K0 Immunoglobulin domain; Lipoprotein; Palmitate; Glycoprotein;
CC K1 Transmembrane; Signal; Antigen.
CC FT CHAIN 1 21
CC FT DOMAIN 22 319 CELL SURFACE A33 ANTIGEN.
CC FT TRANSMEM 22 235 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 236 256 POTENTIAL.
CC FT DOMAIN 257 319 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 22 134 IG-LIKE V-TYPE.
CC FT DOMAIN 140 227 IG-LIKE C2-TYPE.
CC FT DOMAIN 258 261 POLY-CYS.
CC FT DISULFID 43 117 POTENTIAL.
CC FT DISULFID 146 222 POTENTIAL.
CC FT CARBOHYD 162 211 POTENTIAL.
CC FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 319 AA; 35632 MW; 98FC7A4F45C2408E CRC64;

```

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 2 precursor (N-CAM 2).
GN NCAM2 OR NCAM21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=97369930; PubMed=9226371;
RA Paoloni-Giacobino A., Chen H., Antonarakis S.E.;
RT "Cloning of a novel human neural cell adhesion molecule gene (NCAM2)
RT that maps to chromosome region 21q21 and is potentially involved in
RT Down syndrome.";
RL Genomics 43:43-51(1997).
RN [2]
RP CARBOHYDRATE-LINKAGE SITES ASN-445 AND ASN-562.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Abersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003)
CC -1- FUNCTION: May play important roles in selective fasciculation and
CC zone-to-zone projection of the primary olfactory axons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed most strongly in adult and fetal
CC brain.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75330; AAB80803.1; -.
CC DR Genew; HGNC:7657; NCAM2.
CC DR MIM; 602040; -.
CC DR GO; GO:0016021; C:integral to membrane; TAS.
CC DR GO; GO:0005886; C:plasma membrane; TAS.
CC DR GO; GO:0007158; P:neuronal cell adhesion; TAS.
CC DR InterPro; IPR008957; FN III-like.
CC DR InterPro; IPR003961; FN III.
CC DR InterPro; IPR007110; IG_LIKE.
CC DR InterPro; IPR003598; IG_C2.
CC DR Pfam; PF00041; FN3_2.
CC DR SMART; SM0047; IG_5.
CC DR SMART; SM00060; FN3_2.
CC DR SMART; SM00408; IG2_5.
CC DR PROSITE; PSS0835; IG_LIKE; 5.
CC K0 Cell adhesion; Transmembrane; Glycoprotein; Repeat;
CC K1 Immunoglobulin domain; Signal.
CC FT SIGNAL 1 19
CC FT CHAIN 20 837 NEURAL CELL ADHESION MOLECULE 2.
CC FT DOMAIN 20 697 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 698 718 POTENTIAL.
CC FT DOMAIN 719 837 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 21 108 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 113 202 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 208 297 IG-LIKE C2-TYPE 3.
CC FT DOMAIN 302 396 IG-LIKE C2-TYPE 4.
CC FT DOMAIN 401 491 IG-LIKE C2-TYPE 5.
CC FT DOMAIN 482 581 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 594 678 FIBRONECTIN TYPE-III 2.
CC FT DISULFID 42 93 PROBABLE.
CC FT DISULFID 136 186 PROBABLE.
CC FT DISULFID 232 281 PROBABLE.

```

```

FT DISULFID 322 380 PROBABLE.
FT CARBOHYD 422 475
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 837 AA; 92932 MW; C3D034106C5741C1 CRC64;

Query March 8.2%; Score 108; DB 1; Length 837;
Best Local Similarity 19.0%; Pred. No. 1.1;
Matches 45; Conservative 43; Mismatches 85; Indels 64; Gaps 11;

QY 3 RG-VPEHRLVLVQLALLPAATQGNKVVYLGKGDVETELCTASOKKSIOFMKNSNOIKI 61
DB 193 RGEIDFDIYIVNVPATSMQKSFNTAERGEMTSCASGSPFPAISMFRNG-KL 250
QY 62 LGNGSFLTKGPSKLNDAADSRSLMDGNFPLIKNLKIDSDTYICEV-----EDQKE 116
DB 251 IEENEKYLKQ-----SNTLETVRNINSDGGPYVCRAATNKAQEBDEKQ 293
QY 117 E-VOLVFGLTANSPDTHLQ-----GSLVLTLESPPGSSPSVQC----- 155
DB 294 AFLQVFN-----QPHITQLKNETTYNGQ-VTLVCAEGEPIPEITWKAANDGFTTE 345
QY 156 --RSPRG---KNIQGKTLVSQLELDGSGTWTCTVLQ---NOKKVEFKIDIVPR 202
DB 346 GDKSPDRGIEVKGQGGSSSLHKIDVKLGSGRYDEAASRGHQKSYLDIEVAPK 402

RESULT 43
NC22 HUMAN STANDARD; PRT; 761 AA.
ID NC22 HUMAN STANDARD; PRT; 761 AA.
AC P13592; P13593;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM1-120) (CD56 antigen).
GN NCAM1 OR NCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Sketelal muscle;
RA MEDLINE=8305258; PubMed=3253057;
RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA Elsom V., Moore S.E., Goriola C., Walsh F.S.;
RT "Complete sequence and in vitro expression of a tissue-specific
RT phosphatidylinositol-linked N-CAM isoform from sketelal muscle.";
RL Development 104:165-173(1988).
RN [2]
RP SEQUENCE OF 491-761 FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Sketelal muscle;
RA MEDLINE=87301755; PubMed=2887295;
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT "Human muscle neural cell adhesion molecule (N-CAM): identification
RT of a muscle-specific sequence in the extracellular domain.";
RL Cell 50:119-130(1987).
RN [3]
RP SEQUENCE OF 491-655 FROM N.A. (ISOFORM C).
RA MEDLINE=89077552; PubMed=3203385;
RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA Dickson G., Walsh F.S.;
RT "Alternative splicing generates a secreted form of N-CAM in muscle
RT and brain.";
RL Cell 55:955-964(1988).

```

```

CC -1- FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 120;
CC IsoId=P13592-2; Sequence=Displayed;
CC Name=N-CAM 140;
CC IsoId=P13591-1; Sequence=External;
CC Name=C; Synonyms=Secreted;
CC IsoId=P13592-1; Sequence=VSP 002587;
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -1- DATABASE: NMBE=PRO; NOTE=CD guide CD56 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd56.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16841; CA34739.1; -;
DR EMBL; M17409; AA55912.1; -;
DR EMBL; M22094; AA55910.1; -;
DR EMBL; M22092; AA55911.1; -;
DR EMBL; M22091; AA55911.1; JOINED.
DR PIR; A31635; A31635.
DR PIR; S07784; A31635.
DR PIR; S07784; A31635.
DR Genew; HGNC:7656; NCAM1.
DR MIM; 116930; -;
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
DR KEGG; Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW GPI-anchor; Alternative splicing.
FT FT SIGNAL 1 19
FT FT CHAIN 20 761
FT FT
FT FT
FT FT DOMAIN 20 111
FT FT 116 205
FT FT DOMAIN 212 301
FT FT 212 301
FT FT DOMAIN 308 403
FT FT 406 491
FT FT DOMAIN 518 595
FT FT 518 595
FT FT DOMAIN 660 727
FT FT 660 727
FT FT DISULFID 41 96
FT FT 139 189
FT FT DISULFID 235 287
FT FT 329 385
FT FT DISULFID 426 479
FT FT 426 479
FT FT CARBOHYD 222 222
FT FT 315 315
FT FT CARBOHYD 347 347
FT FT 423 423
FT FT CARBOHYD 449 449
FT FT 478 478
FT FT CARBOHYD 498 498
FT FT 635 655
FT FT VARSPLIC
FT FT
FT FT SEQUENCE 761 AA; 83770 MW; F0CAD3292DAB67E CRC64;
SQ

```

```

Query March 8.1%: Score 107; DB 1; Length 761;
Best Local Similarity 19.3%: Pred. No. 1.2;
Matches 44; Conservative 38; Mismatches 102; Indels 44; Gaps 7.

QY 3 RG-VPEFRHLVLVLQALLPAATQGNKVVYLGKKDVTVELTCTASQKKSIOFHKNKNOIKI 61
Db 196 RGEINFKDIOYIVANVPPTIQARONIVANVTANIGQSVTLVCADEGPEEPTMSW----- 247
QY 62 LGNGGSFLTKGPKSKINDADRRSLMDQGNPFLIKNLKIEDSDTYICEVE---DQKEE 117
Db 248 -----TKGGEQIEOEDEDEKRYFSDSSQGLTKVDKNDDEARYICTAENKAGEODAT 299
QY 118 VQLLVFG----LTNSDTHLLQGQSLTTLTLESPSSPVQCR-----SPRKNIQGG- 166
Db 300 IHLKFAKPKTYTYENQTAELIEEYVLTCEASGDPISITWRTISTRINISSEKTLDDGM 359
QY 167 -----KTLVSQLELDQSGTWTG---TVLQNKQKVEFKIDVPR 202
Db 360 VVRSHARVSLTLKSIQYTDAGEYICTASNTIGDQSGSWYLEVQYAPK 407

RESULT 44
NCAL_HUMAN STANDARD: PRT; 848 AA.
ID NCAL_HUMAN
AC P13591; 015829; Q16180:
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neutral cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140) (NCAM1-140) (CD56 antigen).
GN NCAM1 OR NCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94356433; PubMed=8075973;
RA Saito S., Tamio Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT "Complementary DNA sequence encoding the major neural cell adhesion molecule isoform in a human small cell lung cancer cell line.";
RL Lung Cancer 10:307-318(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91250739; PubMed=1710251;
RA Lanier L.B., Chang C., Azuma M., Ruttenberg J.J., Hemperly J.J.,
RA Phillips J.H.;
RT "Molecular and functional analysis of human neural killer cell-associated neural cell adhesion molecule (N-CAM/CD56).";
RL J. Immunol. 146:4421-4426(1991).
[3]
RP SEQUENCE OF 491-848 FROM N.A.
RX MEDLINE=87301755; PubMed=2887295;
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA Moore S.E., Cox R.D., Quinn C., Puet W., Walsh F.S.;
RT "Human muscle neural cell adhesion molecule (N-CAM): identification of a muscle-specific sequence in the extracellular domain.";
RL Cell 50:1119-1130(1987).
-1- FUNCTION: This protein is a cell adhesion molecule involved in neuron-neuron adhesion, neurite fasciculation, outgrowth of neurites, etc.
-1- SUBCELLULAR LOCATION: Type I membrane protein.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=N-CAM 140;
IsoId=PI3591-1; Sequence=Displayed;
Name=N-CAM 120;
IsoId=PI3592-2; Sequence=External;
Name=C; Synonyms=Secreted;
IsoId=PI3592-1; Sequence=External;
-1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
-1- SIMILARITY: Contains 2 fibronectin type III domains.
-1- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;

```

```

WWW="http://www.ncbi.nlm.nih.gov/ncbi/cd/cds5.htm".
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/
CC CC or send an email to license@isb-ib.ch).
-----
DR DR EMBL; S71824; AAB31836.1; -
DR DR EMBL; U63041; AAB04558.1; -
DR DR EMBL; M17410; AAA59913.1; -
DR DR HSSP; P40189; IBOU
DR DR Genew; HGNC;7656; NCAM1.
DR DR MIM; 116930; -
DR DR GO; GO:0016021; C:integral to membrane; TMS.
DR DR GO; GO:0005886; C:plasma membrane; TMS.
DR DR InterPro; IPR008957; FN_III-like.
DR DR InterPro; IPR003961; FN_III.
DR DR InterPro; IPR007110; Ig-like.
DR DR InterPro; IPR003598; Ig_c2.
DR DR Pfam; PF00041; fn3; 2.
DR DR Pfam; PF00047; Ig; 5.
DR DR SMART; SMO0060; FN3; 2.
DR DR SMART; SMO0408; ICG2; 5.
DR DR PROSITE; PS50835; IG_LIKE; 5.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW Transmembrane; Alternative splicing.
FT FT SIGNAL 1 19
FT FT CHAIN 20 848
FT FT ISOCRM. NEURAL CELL ADHESION MOLECULE 1, 140 kDa
FT FT EXTRACELLULAR (POTENTIAL).
FT FT TRANSSEM 709 729
FT FT DOMAIN 730 848
FT FT DOMAIN 20 111
FT FT DOMAIN 116 205
FT FT DOMAIN 212 301
FT FT DOMAIN 308 403
FT FT DOMAIN 406 491
FT FT DOMAIN 518 595
FT FT DOMAIN 660 727
FT FT DISULFID 41 96
FT FT DISULFID 139 189
FT FT DISULFID 235 287
FT FT DISULFID 329 385
FT FT DISULFID 426 479
FT FT CARBOHYD 222 222
FT FT CARBOHYD 315 315
FT FT CARBOHYD 347 347
FT FT CARBOHYD 423 423
FT FT CARBOHYD 449 449
FT FT CARBOHYD 478 478
FT FT CONFLICT 215 215
FT FT CONFLICT 239 239
FT FT CONFLICT 490 490
FT FT CONFLICT 599 600
FT FT CONFLICT 720 721
FT FT CONFLICT 811 811
SQ SEQUENCE 848 AA; 93360 MW; 66D2P0C0E6C1C2AD CRC64;
Query Match 8.1%; Score 107; DB 1; Length 848;
Best Local Similarity 19.3%; Pred. No. 1.4;
Matches 44; Conservative 38; Mismatches 102; Indels 44; Gaps 7;
-----
QY 3 RG-VPEFRLILVLDLALPAAQTGNKVKLVKKGGDTVLTCTAAGKSKSIQFHWKNSNQDIKI 61
DB 196 RGEINFKDIOIYIVNVPPTIQARQIVVATNANLGQSVTLVCDAGCFPEPTMSW----- 247
QY 62 LGNGSGFLTKPSKLNDRADRSRLMPCGNPFLIKIKMLKEDSDTYICEVE-----DQKEE 117
DB 248 -----TDGGEQIEQEBDEDEKRIFFSDSSQLTIKVKVDKNDLAEFYICIAENKAGQDAT 299

```

```

Oy 118 VOLVFG---LTANSPTHLLQSGSLTTLSPSSPSVOCR-----SPRGKNIQCG- 166
Dy 300 IHLKVFAPKITYVENQFAMEBEVYTLTCEASGPISITMRTSTRNISSEKTLDDHM 359
Dy 167 -----KTLVSQLELDGSGTWTCT---TVLQNKVFEKIDIVP 202
Dy 360 VVRSHARVSLTSLKSIQYTDAGEYICTASNTIGQSGMYLEVQYAPK 407

RESULT 45
DCC_MOUSE STANDARD; PRT; 1447 AA.
ID DCC_MOUSE
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tumor suppressor protein DCC precursor.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=96112625; Pubmed=8570174;
RA Cooper H.M., Armes P., Brito J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
RT gene (mdc) and its expression in the developing mouse embryo.";
RL Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=P70211-1; Sequence=Displayed;
CC Note=Isoform B is produced by alternative initiation at Met-85
CC of isoform A;
CC Name=C;
CC IsoId=P70211-2; Sequence=VSP_002501;
CC Event=Alternative initiation;
CC Comment=2 isoforms, A (shown here) and B, are produced by
CC alternative initiation at Met-1 and Met-85;
CC -1- TISSUE SPECIFICITY: In the embryo, expressed at high levels in the
CC developing brain and neural tube. In adult, highly expressed in
CC brain with very low levels found in testis, heart and thymus.
CC Isoform C is expressed only in the embryo.
CC -1- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels
CC expressed during mid gestation. Levels decrease in late gestation
CC and remain at this level in the adult.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X85788; CA59786.1; -.
CC HSP; P56276; IRLK.
CC MGI; MGI:94869; DCC.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FN_III_subd.

```

```

DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; ENTPEIIT.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00408; IGc2; 3.
DR PROSITE; PS00835; IG_LIKE; 4.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
KW SIGNAL
FT SIGNAL 1 25
FT CHAIN 26 1447
FT CHAIN 85 1447
FT INT MET 85 85
FT DOMAIN 26 1097
FT TRANSMEM 1098 1122
FT DOMAIN 1123 1447
FT DOMAIN 36 135
FT DOMAIN 139 229
FT DOMAIN 234 326
FT DOMAIN 331 416
FT DOMAIN 426 522
FT DOMAIN 525 618
FT DOMAIN 619 716
FT DOMAIN 722 816
FT DOMAIN 840 940
FT DOMAIN 941 1042
FT DISULFID 61 117
FT DISULFID 161 212
FT DISULFID 261 310
FT DISULFID 352 400
FT CARBOHYD 60 60
FT CARBOHYD 94 94
FT CARBOHYD 299 299
FT CARBOHYD 318 318
FT CARBOHYD 478 478
FT CARBOHYD 628 628
FT CARBOHYD 702 702
FT VASPLIC 819 838
SQ SEQUENCE 1447 AA; 158298 MW; 0D1F1097C2D5B9F CRC64;

Query Match 8.1%; Score 107; DB 1; Length 1447;
Best Local Similarity 24.1%; Pred. No. 2; 7;
Matches 46; Conservative 26; Mismatches 75; Indels 44; Gaps 8;

Oy 34 GDTVELTCTASQKKSIOFHMKNSNQIKLGNQGSFLTKGPSKLD-RADSRRLMDQGNF 92
Dy 154 GDTVLACVIGBEMPTIIQKNGQ-----DLNPLPGDSRVVVLPGS-- 195
Oy 93 PLIKNLKIEDSDTYICEVED-----QKEEVOLL-----VFGLTANSPTHLLQCG 137
Dy 196 ALQISRLQPGSGGYRCASNPASIRTNEMAEVAILDPGLHROLVFLQRPNSVIALEGK 255
Oy 138 SLTL---TLSPSS-----SPSVQGRSPRGKNIQSGKTLVSQLELDGSGTWTCTVYLQ 188
Dy 256 DAIVLCCVGGVPPSPFTWLAGBEVYQLRSKK-YSLDGSNLLISNVTDSDSGITTCVYTY 314
Oy 189 NQKVEFEKIDI 199
Dy 315 KNEIISASDEL 325

RESULT 46
PGBM_MOUSE STANDARD; PRT; 3707 AA.
ID PGBM_MOUSE
AC O05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PVC).
GN HSPG2.

```

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Scleroglossa; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=92078153; PubMed=1744087;  
 RA Noonan D.M., Fuller A., Valente P., Cai S., Horigan E., Sasaki M.,  
 RA Yamada Y., Hassell J.R.;  
 RT "The complete sequence of perlecan, a basement membrane heparan  
 sulfate proteoglycan, reveals extensive similarity with laminin A  
 chain, low density lipoprotein receptor, and the neural cell adhesion  
 molecule";  
 RL J. Biol. Chem. 266:22939-22947 (1991).  
 [2]  
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89034110; PubMed=2972708;  
 RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,  
 RA Yamada Y., Hassell J.R.;  
 RT "Identification of cDNA clones encoding different domains of the  
 basement membrane heparan sulfate proteoglycan";  
 RL J. Biol. Chem. 263:16379-16387 (1988).  
 CC -1- FUNCTION: This protein is an integral component of basement  
 CC membranes. It is responsible for the fixed negative electrostatic  
 CC charge and is involved in the charge-selective ultrafiltration  
 CC properties. It serves as an attachment substrate for cells.  
 CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in  
 CC dimers or stellate structures. It interacts with other basement  
 CC membrane components such as laminin, prolargin and collagen type  
 CC IV.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Found in the basement membranes.  
 CC -1- PFM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED  
 CC AND O-LINKED OLIGOSACCHARIDES.  
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
 CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.  
 CC -1- SIMILARITY: Contains 3 laminin IV domains.  
 CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 3 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 EGF-like domain.  
 CC -1- SIMILARITY: Contains 1 SEA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; M77174; AAA39911.1; -;  
 DR EMBL; J04054; AAA39899.1; -;  
 DR EMBL; J04055; AAA39912.1; -;  
 DR PIR; S18252; S18252.  
 DR PDB; 1GLA; 28-NOV-01.  
 DR MGD; MGI:96257; Hsp92.  
 DR GO; GO:0005604; C:basement membrane; IDA.  
 DR GO; GO:0008104; P:protein localization; IMP.  
 DR InterPro; IPR008985; ConA\_like\_1ec\_g1.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR InterPro; IPR000034; laminin\_B.  
 DR InterPro; IPR002049; laminin\_EGF.  
 DR InterPro; IPR001791; laminin\_G.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF00008; EGF; 4.  
 DR Pfam; PF00047; IG; 15.  
 DR Pfam; PF00053; laminin\_B; 3.  
 DR Pfam; PF00053; laminin\_EGF; 7.

DR Pfam; PF00054; laminin\_G; 3.  
 DR Pfam; PF00057; LDL\_recept\_a; 4.  
 DR Pfam; PF01390; SEA; 1.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PRODOM; PD003031; laminin\_B; 3.  
 DR SMART; SM00180; EGF\_Lam; 7.  
 DR SMART; SM00408; IGC2; 14.  
 DR SMART; SM00281; Lamb; 3.  
 DR SMART; SM00282; Lamb; 3.  
 DR SMART; SM00192; LDLa; 4.  
 DR SMART; SM00200; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 8.  
 DR PROSITE; PS01186; EGF\_2; 5.  
 DR PROSITE; PS00026; EGF\_3; 4.  
 DR PROSITE; PS00835; IG\_LIKE; 15.  
 DR PROSITE; PS00825; ILM\_G\_DOMAIN; 3.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
 DR PROSITE; PS01209; LDLRA\_1; 4.  
 DR PROSITE; PS00068; LDLRA\_2; 4.  
 DR PROSITE; PS00024; SEA; 1.  
 DR SIGNAL; Basement membrane; Proteoglycan; Repeat; Glycoprotein;  
 DR Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;  
 DR Extracellular matrix; EGF-like domain; 3D-structure.  
 FT CHAIN 1 21  
 FT SIGNAL 1 21  
 FT 22 3707  
 FT 80 194  
 FT 195 234  
 FT 281 319  
 FT 320 359  
 FT 360 403  
 FT 404 504  
 FT 521 530  
 FT 531 730  
 FT 731 763  
 FT 764 813  
 FT 814 871  
 FT 879 923  
 FT 924 933  
 FT 934 1125  
 FT 1126 1158  
 FT 1159 1208  
 FT 1209 1265  
 FT 1275 1324  
 FT 1325 1334  
 FT 1335 1529  
 FT 1530 1562  
 FT 1563 1612  
 FT 1613 1670  
 FT 1677 1771  
 FT 1772 1865  
 FT 1866 1954  
 FT 1955 2049  
 FT 2049 2148  
 FT 2149 2244  
 FT 2245 2343  
 FT 2344 2436  
 FT 2437 2532  
 FT 2533 2619  
 FT 2620 2720  
 FT 2721 2809  
 FT 2810 2895  
 FT 2896 2980  
 FT 2984 3162  
 FT 3163 3241  
 FT 3245 3425  
 FT 3518 3705  
 FT 3518 3705  
 FT SITE 65 67  
 FT SITE 71 73  
 FT SITE 76 78  
 FT SITE 3615 3617  
 FT DISULFID 199 212  
 BRASEMENT MEMBRANE-SPECIFIC HEPARAN  
 SULFATE PROTEOGLYCAN CORE PROTEIN.  
 SEA.  
 LDL-RECEPTOR CLASS A 1.  
 LDL-RECEPTOR CLASS A 2.  
 LDL-RECEPTOR CLASS A 3.  
 LDL-RECEPTOR CLASS A 4.  
 IG-LIKE C2-TYPE 1.  
 LAMININ EGF-LIKE 1 (N-TERMINAL).  
 LAMININ DOMAIN IV 1 (DOMAIN III A).  
 LAMININ EGF-LIKE 1 (C-TERMINAL).  
 LAMININ EGF-LIKE 2.  
 LAMININ EGF-LIKE 3.  
 LAMININ EGF-LIKE 4 (INCOMPLETE).  
 LAMININ EGF-LIKE 5 (N-TERMINAL).  
 LAMININ DOMAIN IV 2 (DOMAIN III B).  
 LAMININ EGF-LIKE 5 (C-TERMINAL).  
 LAMININ EGF-LIKE 6.  
 LAMININ EGF-LIKE 7.  
 LAMININ EGF-LIKE 8.  
 LAMININ EGF-LIKE 9 (N-TERMINAL).  
 LAMININ DOMAIN IV 3 (DOMAIN III C).  
 LAMININ EGF-LIKE 9 (C-TERMINAL).  
 LAMININ EGF-LIKE 10.  
 LAMININ EGF-LIKE 11.  
 IG-LIKE C2-TYPE 2.  
 IG-LIKE C2-TYPE 3.  
 IG-LIKE C2-TYPE 4.  
 IG-LIKE C2-TYPE 5.  
 IG-LIKE C2-TYPE 6.  
 IG-LIKE C2-TYPE 7.  
 IG-LIKE C2-TYPE 8.  
 IG-LIKE C2-TYPE 9.  
 IG-LIKE C2-TYPE 10.  
 IG-LIKE C2-TYPE 11.  
 IG-LIKE C2-TYPE 12.  
 IG-LIKE C2-TYPE 13.  
 IG-LIKE C2-TYPE 14.  
 IG-LIKE C2-TYPE 15.  
 LAMININ G-LIKE 1.  
 EGF-LIKE.  
 LAMININ G-LIKE 2.  
 LAMININ G-LIKE 3.  
 HEPARAN SULFATE (POTENTIAL).  
 HEPARAN SULFATE (POTENTIAL).  
 HEPARAN SULFATE (POTENTIAL).  
 MEDIATES MOTOR NEURON ATTACHMENT  
 (POTENTIAL).  
 BY SIMILARITY.



```

FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 428 479 BY SIMILARITY.
FT DISULFID 764 773 BY SIMILARITY.
FT DISULFID 766 780 BY SIMILARITY.
FT DISULFID 783 792 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 814 829 BY SIMILARITY.
FT DISULFID 816 839 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 869 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
FT DISULFID 1161 1175 BY SIMILARITY.
FT DISULFID 1178 1187 BY SIMILARITY.
FT DISULFID 1190 1206 BY SIMILARITY.
FT DISULFID 1209 1224 BY SIMILARITY.
FT DISULFID 1211 1234 BY SIMILARITY.
FT DISULFID 1237 1246 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1275 1287 BY SIMILARITY.
FT DISULFID 1275 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1307 1322 BY SIMILARITY.
FT DISULFID 1563 1572 BY SIMILARITY.
FT DISULFID 1565 1579 BY SIMILARITY.
FT DISULFID 1582 1591 BY SIMILARITY.
FT DISULFID 1594 1610 BY SIMILARITY.
FT DISULFID 1613 1628 BY SIMILARITY.
FT DISULFID 1615 1638 BY SIMILARITY.
FT DISULFID 1641 1650 BY SIMILARITY.
FT DISULFID 1653 1668 BY SIMILARITY.
FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 8.1%; Score 107; DB 1; Length 3707;
Best Local Similarity 21.7%; Pred. No. 8.5;
Matches 55; Conservative 39; Mismatches 95; Indels 64; Gaps 10;

```

```

OPCM RAT
ID OPCM RAT STANDARD; PRT; 345 AA.
AC P32735; P32735; Q01653; Q01654;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE (Opioid-binding cell adhesion molecule) (OPCM).
GN OPCML OR OBCAM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92347701; Pubmed=1339369;
RA Lippman D.A., Lee N.M., Loh H.H.;
RT "Opioid-binding cell adhesion molecule (OBCAM)-related clones from a
RT rat brain cDNA library.";
RL Gene 117:249-254(1992).
RN [2]
RP SEQUENCE OF 195-214, AND GPI-ANCHOR.
RX MEDLINE=95198094; Pubmed=7891157;
RA Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA Salzer J.L.;
RT "Cloning of neurotrophin defines a new subfamily of differentially
RT expressed neutral cell adhesion molecules.";
RL J. Neurosci. 15:2141-2156(1995).
CC -1- FUNCTION: Binds opioids in the presence of acidic lipids; probably
CC involved in cell contact.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P32736-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P32736-2; Sequence=VSP_002612;
CC Note=No experimental confirmation available;
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M88710; AAA40859.1; -
DR EMBL; M88711; AAA40860.1; -
DR EMBL; M88709; AAA40858.1; -
DR PIR; JCI238; JCI238.
DR PIR; JCI239; JCI239.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IgC2; 2.
DR PROSITE; PSS0835; IG_LIKE; 3.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW Repeat; Signal; Alternative splicing; Lipoprotein.
FT SIGNAL 1 27
FT CHAIN 28 322
FT PROPEP 323 345
FT DOMAIN 39 126
FT DOMAIN 136 219
FT DOMAIN 223 310
FT DISULFID 57 115
FT DISULFID 157 202

```

```
FT DISULFID 244 296 POTENTIAL.
FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 322 322 GPI-anchor amidated asparagine
(Potential).
FT VARSPLIC 1 27 MGVCGLFLPWLKLVVSLFLPVT -> MYHPAYWIF
SATVLFIP (in isoform 2).
FT FT /FtId=VSP_002612.
SQ SEQUENCE 345 AA; 38067 MW; A3181B0753F9658E CRC64;
Query Match 8.1%; Score 106.5; DB 1; Length 345;
Best Local Similarity 22.1%; Pred. No. 0.5;
Matches 58; Conservative 43; Mismatches 101; Indels 61; Gaps 10;
QY 11 LVVLQALALPAALPQGNKVVIG-----KKGDTVELPCTASQKKSIGFMKNSN 57
DB 14 LVVSLKLLFLVPTGVPVRSQDATFPKAMDNVTVRQGSATLCTDDRYT-KVAMLNRS 72
QY 58 QIKILGN-----QGSFLTGPSKLNDRASRSRLWDQNPFLIKNLKEDSDTYICE 110
DB 73 TLIVAGDKMSIDPRVILVNTPTQ-----YSIMIQNVVDEGPYTCS 116
QY 111 VE---DQKEEVOLLVFG---LTNSDTHLLOGSLTLTLESPGSSPSVQCRSPRGKN 162
DB 117 VQVDNPKTSRVHLIVVPPQIMNISSDITVNEISSVTLCLAIQREPFVTRHLSVKE 176
QY 163 IGG---GKTLVSQLELDQSGMTCTVQLQ-----NOKKVEFIDIVPRASALPAPPTS 213
DB 177 GGGVSEDEYLESDIRDSQGEYECALNDVAAPDKVKITVNPPTYS--KAKXGV 234
QY 214 ALPDPQTASALPDPASALPAA 236
DB 235 SVGQKGLSC-----EASAVPMA 252
RESULT 48
DSCA HUMAN STANDARD; PRT; 2012 AA.
ID DSCA HUMAN 060469;
AC 060469; 060468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=98087574; Pubmed=9426258;
RA Yamakawa K., Huoc Y.-K., Haendel M.A., Hubert R., Chen X.-N.,
RA Lyons G.B., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system";
RT Hum. Mol. Genet. 7:227-237(1998).
RL [2]
RN SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20384934; Pubmed=10925149;
RA Aarawala K.L., Nakamura S., Teutsuni Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN [3]
RX SEQUENCE FROM N.A.
RP MEDLINE=20289799; Pubmed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
```

```
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Aakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuoka S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Kameier J., Beck A., Klages S., Hennig S., Rieseemann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=CHD2-52;
CC IsoId=O60469-1; Sequence=Displayed;
CC Name=Short; Synonyms=CHD2-42;
CC IsoId=O60469-2; Sequence=VSP_002502, VSP_002503;
CC -1- TISSUE SPECIFICITY: Primarily expressed in brain.
CC -1- SIMILARITY: Contains 10 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL database.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF023450; AAC17967.1; -
CC EMBL: AF023449; AAC17966.1; -
CC EMBL: AF217525; AAF27525.1; -
CC EMBL: AL163283; CAB90464.1; -
CC EMBL: AL163282; CAB90436.1; -
CC EMBL: AL163281; CAB90444.1; -
CC Genew: HGNC:3039; DSCAM.
CC MIM: 602523; -
CC DR GO: 0005887; C: integral to plasma membrane; TAS.
CC DR GO: 0005624; C: membrane fraction; TAS.
CC DR GO: 0007155; P: cell adhesion; TAS.
CC DR GO: 0007399; P: neurogenesis; TAS.
CC DR InterPro: IPR008957; FN III-like.
CC DR InterPro: IPR003961; FN III.
CC DR InterPro: IPR007110; IG-like.
CC DR InterPro: IPR003598; IG_C2.
CC DR Pfam: PF00041; fn3; 6.
CC DR Pfam: PF00047; Ig; 9.
CC DR SMART: SM00060; FN3; 6.
CC DR SMART: SM00408; IGC2; 7.
CC DR PROSITE: PS00835; IG_LIKE; 9.
CC DR Transmembrane domain; Glycoprotein; Signal; Cell adhesion; Repeat;
CC Immunoglobulin domain; Alternative splicing.
CC KW Transmembrane; Alternative splicing.
CC FT SIGNAL 1 17
CC FT CHAIN 18 2012
CC FT DOMAIN 18 1595
CC FT TRANSMEM 1596 1616
CC FT DOMAIN 1617 2012
CC FT DOMAIN 39 129
CC FT DOMAIN 125 216
CC FT DOMAIN 225 305
CC FT DOMAIN 313 401
CC FT DOMAIN 407 500
CC FT DOMAIN 504 592
CC FT DOMAIN 596 685
CC FT DOMAIN 690 783
CC FT DOMAIN 787 883
CC IG-LIKE C2-TYPE 1.
CC IG-LIKE C2-TYPE 2.
CC IG-LIKE C2-TYPE 3.
CC IG-LIKE C2-TYPE 4.
CC IG-LIKE C2-TYPE 5.
CC IG-LIKE C2-TYPE 6.
CC IG-LIKE C2-TYPE 7.
CC IG-LIKE C2-TYPE 8.
CC IG-LIKE C2-TYPE 9.
```

```

FT DOMAIN 885 972 FIBONECTIN TYPE-III 1.
FT DOMAIN 984 1076 FIBONECTIN TYPE-III 2.
FT DOMAIN 1088 1177 FIBONECTIN TYPE-III 3.
FT DOMAIN 1189 1273 FIBONECTIN TYPE-III 4.
FT DOMAIN 1285 1377 IG-LIKE C2-TYPE 10.
FT DOMAIN 1380 1463 FIBONECTIN TYPE-III 5.
FT DOMAIN 1477 1562 FIBONECTIN TYPE-III 6.
FT DISULFID 46 102 BY SIMILARITY.
FT DISULFID 145 197 BY SIMILARITY.
FT DISULFID 246 293 BY SIMILARITY.
FT DISULFID 335 385 BY SIMILARITY.
FT DISULFID 428 484 BY SIMILARITY.
FT DISULFID 525 575 BY SIMILARITY.
FT DISULFID 617 669 BY SIMILARITY.
FT DISULFID 711 766 BY SIMILARITY.
FT DISULFID 809 865 BY SIMILARITY.
FT DISULFID 1307 1359 BY SIMILARITY.
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 795 795 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1250 1250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1271 1271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1488 1488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASPLIC 1562 1571 NFATL1NDGS -> KEARCKEFS (in isoform short).
FT VASPLIC 1572 2012 Missing (in isoform short).
FT CONFLICT 1893 2012 /FTid=VSP_002503.
FT SRTLRPTVLEPIRMEASASSTRREGSQMGAVATLPOR
FT EGAEIGQAKKSSQESLDRGLKGNKPNYAKSYTLV ->
FT IGVSTYCIHTEMTFC (IN REF. 1).
SQ SEQUENCE 2012 AA; 222259 MM; 0E3JCEB781A0834 CRC64;

Query Match 8.0%; Score 106; DB 1; Length 2012;
Best Local Similarity 21.5%; Pred. No. 4.8;
Matches 50; Conservative 34; Mismatches 75; Indels 74; Gaps 10;

OY 34 GDTVELTCTASQKSIQPHWKNQIKLGNQSFLLTKGPKLMDRADRSRLMDQGNFP 93
DB 239 GQVLELPCALGHEPDRMLKDNM-----PLELSGR-----FQKVTG 277
OY 94 LIINKLEBDSYICEVEDQKEVQLLVFGLTANSDTHLLOQSLTTL-----ESPFG 148
DB 278 LILININPSDGSSTVEVSNR-----YGTAKVTIGRLVVKPLAKTISPRVKSSVG 328
OY 149 SSPSVOCR-----SPRGKNIQ-----GKTLTSLVSOLELDQSGTWTCTV 186
DB 329 SSVLSSTVGTEDEQLSWYNGEILNP-GSNVRIITGINHMLMDHNVVSDGAYQCFV 387
OY 187 LQNKVVEFKIDIVPRASALPAPPTGSALPD--PQNASALDPDPAASALPAL 237
DB 388 RKDKLSADYVQV-----LEDGTETKIIISAFSEKVSAPAEVSL 426

RESULT 49
NTRI_HUMAN STANDARD; PRT; 344 AA.
AC O9P121;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin precursor (hnt).
GN NT...
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li G., Jin J., Tan X., Hu S., Yuan J., Qiang B.;
RT "Cloning and identification of human neurotrophin full length cDNA.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Neural cell adhesion molecule.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF126426; AAF37591.1; -.
DR MIM; 607938; -.
DR GO; GO:0008038; P:neural cell recognition; TAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IgC2; 2.
DR PROSITE; PS50835; IG-LIKE; 3.
DR KX Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KM Repeat; Signal; Lipoprotein.
FT CHAIN 1 31 POTENTIAL.
FT PROPEP 32 321 NEUROTROPHIN.
FT DOMAIN 322 344 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 339 126 IG-LIKE C2-TYPE 1.
FT DOMAIN 136 218 IG-LIKE C2-TYPE 2.
FT DOMAIN 222 309 IG-LIKE C2-TYPE 3.
FT DISULFID 57 115 POTENTIAL.
FT DISULFID 157 201 POTENTIAL.
FT DISULFID 243 295 POTENTIAL.
FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 321 321 GPI-anchor amidated asparagine
(Potential).
SQ SEQUENCE 344 AA; 37971 MM; DA4D12C295ABBEBA CRC64;

Query Match 8.0%; Score 105; DB 1; Length 344;
Best Local Similarity 20.8%; Pred. No. 0.65;
Matches 53; Conservative 48; Mismatches 108; Indels 46; Gaps 9;

OY 11 LVVLQALLPLRAAOGKRVVG-----KKGDTVELTCTASQKSIQPHWKN 57
DB 14 LVVLSLELLPLVFGVVRSGDATFPKAMDNVTVRQESATLCTTDNRVT-RVAMINRS 72
OY 58 QIKLGNQSFLLTKGPKLMDRADRSRLMDQGNFPLIINKLEBDSYICEVEDQKEE 117
DB 73 TILVAGDKKVCIDPRVVLN-----TQYQYSEIIONVYVDEGPTTCSVQDNHP 123
OY 118 -----VQLLVFGLTANSDTHLLOQSLTTLSPGSSPSVOCR--SPRGKN-IOGG 166
DB 124 KTSRVHLIVQVSPKIVAISSDISINEGNISTLCIATGPREPTVTRHISPKAVGVSED 183
OY 167 KTLTSLVSOLELDQSGTWTCTVLQN-----QKKVEFKIDIVPRASALPAPPTGSA 221

```

```

Db      184 EYEIIGITREOSGDECSASNDVAPVVRKVTNVPYIS----EAKGTGVPVGQKG 239
QY      222 SALPPPPASALPAA 236
Db      240 TLQCE---ASAVPSA 251

RESULT 50
NTRI_MOUSE STANDARD; PRT; 344 AA.
AC 099PJO;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.C.,
RA Rana S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKeown P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rutterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Neural cell adhesion molecule.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF282980; AAK00276.1; -
CC EMBL; BC023307; AAH23307.1; -
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003598; Ig_c2.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00409; IG_3.
CC SMART; SM00408; IGC2; 3.

```

```

DR PROSITE; PSS0835; IG LIKE; 3.
KW Immunoglobulin domain, Cell adhesion, Glycoprotein, GPI-anchor;
KW Repeat; Signal; Lipoprotein.
FT SIGNAL 1 31
FT CHAIN 32 321
FT PROPEP 322 344
FT DOMAIN 39 126
FT DOMAIN 136 218
FT DOMAIN 222 309
FT DISULFID 57 115
FT DISULFID 157 201
FT DISULFID 243 295
FT CARBOHYD 44 44
FT CARBOHYD 70 70
FT CARBOHYD 152 152
FT CARBOHYD 284 284
FT CARBOHYD 292 292
FT CARBOHYD 305 305
FT CARBOHYD 321 321
FT LIPID 321 321
FT CONFLICT 75 75
FT CONFLICT 92 92
FT CONFLICT 119 119
FT CONFLICT 187 187
FT CONFLICT 213 213
FT CONFLICT 225 225
SQ SEQUENCE 344 AA; 37984 MW; C885BBA52C148554 CRC64;

Query Match 8.0%; Score 105; DB 1; Length 344;
Best Local Similarity 20.8%; Pred. No. 0.65;
Matches 53; Conservative 48; Mismatches 108; Indels 46; Gaps 9;

QY 11 LVLQLALIPAAATQGNKVLG-----KKGDTVELTCTASQKKSIOFHKNSN 57
Db 14 LVVSLRLFLVPTGPVPSGDATEPKAMDNTVVRGSESATRCITDNRVT-RVAMLNIS 72
QY 58 QKIIGNOSPLTKGSKINDRADRSRLMDGNFLLIKNIKIEDSDTYICEVEQKES 117
Db 73 TLVYANDKRCCLDPVRLVLSN-----TQVSIIEQVNDVDEGPTYSVQTDNHP 123
QY 118 -----VQLVFGLTANSDTHLLQGSILTTLESPPSSPSVOCR--SPRKN-IOGG 166
Db 124 KTSRVALLVQVBPKEIVISDSISNGNNSILTCATGPEPTVTRHISPKAVGVSD 183
QY 167 KTLVSQGLQDSGTCTVLQN-----QKXVEFKDIYPRASALPAPRPGSALPPQRTA 221
Db 184 EYEIIGITREOSGDECSASNDVAPVVRKVTNVPYIS----EAKGTGVPVGQKG 239
QY 222 SALPPPPASALPAA 236
Db 240 TLQCE---ASAVPSA 251

RESULT 51
NTRI_MOUSE STANDARD; PRT; 344 AA.
AC Q62718;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin precursor (GP65).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 217-229.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=9518094; PubMed=7891157;
RA Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA Salzer J.L.;
RT "Cloning of neurotrophin defines a new subfamily of differentially

```

RT expressed neural cell adhesion molecules.";  
 RL J. Neurosci. 15:2141-2156(1995).  
 CC -1- FUNCTION: Neural cell adhesion molecule.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: Central nervous system.  
 CC -1- DEVELOPMENTAL STAGE: Expressed at high levels in several  
 CC subplate, and lower cortical laminae in the forebrain, and in the  
 CC portine nucleus, cerebellar granule cells, and Purkinje cells in  
 CC the hindbrain.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLN  
 CC family.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U16845; AAB67445.1; -.  
 CC DR PIR; I56551; I56551.  
 CC DR InterPro; IPR007110; IG\_1like.  
 CC DR InterPro; IPR003598; IG\_c2.  
 CC DR Pfam; PF00047; Ig\_3.  
 CC DR SMART; SM00408; IGc2; 2.  
 CC DR PROSITE; PS50835; IG\_LIKE; 3.  
 CC KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 CC Repeat; Signal; Lipoprotein.  
 FT SIGNAL 1 31  
 FT CHAIN 32 321  
 FT PROPEP 322 344  
 FT DOMAIN 39 126  
 FT DOMAIN 136 218  
 FT DOMAIN 222 309  
 FT DISULFID 57 115  
 FT DISULFID 157 201  
 FT DISULFID 243 295  
 FT CARBOHYD 44 70  
 FT CARBOHYD 152 152  
 FT CARBOHYD 216 216  
 FT CARBOHYD 284 284  
 FT CARBOHYD 292 292  
 FT CARBOHYD 305 305  
 FT CARBOHYD 321 321  
 FT LIPID 321 321  
 SQ SEQUENCE 344 AA; 37998 MW; CB839BE5B3B224 CRC64;  
 Query Match 7.9%; Score 104; DB 1; Length 344;  
 Best Local Similarity 20.8%; Pred. NO. 0.77;  
 Matches 53; Conservative 47; Mismatches 109; Indels 46; Gaps 9;  
 Oy 11 LVLVQLALLPAAIGKQKVVLC-----KKGDTVELTCTAQSCKSIQTHWKNXN 57  
 Db 14 LVVVSLRLFLVPTGVPRVSGDAPFPKAMDVTVVAGSGSATLRCTIDNRVT-RVAMLNRS 72  
 Oy 58 QIKLNGSGFLTKGPKSLNDRBSRLMDQGNPLIKLKIEDSTYICEVEDQKEE 117  
 Db 73 TILYAGNDKWCIDPRVVLN-----TOTOISIEIQNVADVBSGPTCSVOTDNP 123  
 Oy 118 -----VQLLVFGLTANSPTHLTQSGSLTTLSPGSSPSVQCR--SPRGK-IOGG 166  
 Db 124 KTSRVHLIVQSPKIVEISDISINISGTCATGRPEPTVWRIISFKAVGFSVED 183  
 Oy 167 KTLVSQLELDGSGTWCTVLQN-----QKVEFKIDIVPAPASALPAPFGSALPDPQTA 221  
 Db 184 EYLEIOGTIRBOGSEYECSSANDVAAPVVRVNTVNPYPYIS-----EAKGTGVVQKG 239  
 Oy 222 SALDPPAPASALPAA 236

Db 240 TLQCE---ASAVPSA 251  
 RESULT 52  
 CONT HUMAN STANDARD; PRT; 1018 AA.  
 ID CONT HUMAN 012860; 014030;  
 AC 012860; 012861; 014030;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Contactin precursor (Glycoprotein gp135).  
 GN CNTN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=95048335; PubMed=7959734;  
 RA Berglund E.O., Kanacht B.;  
 RT "Molecular cloning and in situ localization of the human contactin  
 RT gene (CNTN1) on chromosome 12q11-q12.";  
 RL Genomics 21:571-582(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=94217459; PubMed=8164510;  
 RA Reid R.A., Hemperly J.J.;  
 RT "Identification and characterization of the human cell adhesion  
 RT molecule contactin.";  
 RL Brain Res. Mol. Brain Res. 21:1-8(1994).  
 CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS SYSTEM  
 CC DEVELOPMENT. IN ASSOCIATION WITH CNTNAP1 SEEMS TO PLAY A ROLE IN  
 CC THE FORMATION OF PARANODAL AXO-GLIAL JUNCTIONS IN MYELINATED  
 CC PERIPHERAL NERVES AND MAY HAVE A ROLE IN THE SIGNALING BETWEEN  
 CC AXONS AND MYELINATING GLIAL CELLS.  
 CC -1- SUBUNIT: INTERACTS WITH CONTACTIN ASSOCIATED PROTEIN 1 IN CIS  
 CC FORM. BINDS TO THE CARBONIC-ANHYDRASE LIKE DOMAIN OF PROTEIN-  
 CC TYROSINE PHOSPHATASE ZETA (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q12860-1; Sequence=displayed;  
 CC Name=2;  
 CC IsoId=Q12860-2; Sequence=VSP\_002500;  
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 4 fibronectin type III domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U07819; AAB67920.1; -.  
 CC DR EMBL; U07820; AAB67921.1; -.  
 CC DR EMBL; Z21488; CAA79696.1; -.  
 CC DR PIR; A54744; A54744.  
 CC DR HSSP; P08921; IAB7.  
 CC DR Genew; HGNC:2171; CNTN1.  
 CC DR MIM; 600016; -.  
 CC GO; GO:0005624; C:membrane fraction; TAS.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR Pfam; PF00041; fn3; 4.  
 DR Pfam; PF00047; Ig; 6.  
 DR SMART; SM00600; FN3; 4.

DR SMART: SM00408; IGC2; 4.  
 DR PROSITE; PS50835; IG\_LIKE; 6.  
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;  
 Cell adhesion; Repeat; Alternative splicing; Polymorphism;  
 KM Lipoprotein.

FT	CHAIN	1	20	CONTACTIN.
FT	PROPER	21	2993	REMOVED IN MATURE FORM.
FT	DOMAIN	7994	1018	IG-LIKE C2-TYPE 1.
FT	DOMAIN	41	131	IG-LIKE C2-TYPE 2.
FT	DOMAIN	137	223	IG-LIKE C2-TYPE 3.
FT	DOMAIN	241	326	IG-LIKE C2-TYPE 4.
FT	DOMAIN	331	407	IG-LIKE C2-TYPE 5.
FT	DOMAIN	413	500	IG-LIKE C2-TYPE 6.
FT	DOMAIN	504	601	GLY/PRO-RICH.
FT	DOMAIN	602	609	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	609	710	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	711	812	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	813	908	FIBRONECTIN TYPE-III 4.
FT	DOMAIN	909	1004	FIBRONECTIN TYPE-III 5.
FT	DISULFID	65	114	BY SIMILARITY.
FT	DISULFID	158	211	BY SIMILARITY.
FT	DISULFID	263	310	BY SIMILARITY.
FT	DISULFID	352	391	BY SIMILARITY.
FT	DISULFID	436	484	BY SIMILARITY.
FT	CARBOHYD	208	208	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	258	258	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	494	494	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	521	521	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	591	591	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	933	933	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	LIPID	993	993	GPI-anchor amidated serine (Potential).
FT	VARSPLIC	21	31	Missing (in isoform 2).
FT	VARIANT	798	798	V->L (in dbSNP:1056020).
FT	SEQUENCE	1018 AA; 113320 MW; 488PDCSBDP434ED5 CRC64;		/FTID=VAR_011722.
Query Match	Best local Similarity	7.9%; Score 104; DB 1; Length 1018;		
Matches	51; Conservative	27; Mismatches	90; Indels	82; Gaps

DR SMART: SM00408; IGC2; 4.  
 DR PROSITE; PS50835; IG\_LIKE; 6.  
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;  
 Cell adhesion; Repeat; Alternative splicing; Polymorphism;  
 KM Lipoprotein.

DE Contactin precursor (Neural cell surface protein F3).  
 GN CTNNA1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA MEDLINE=89340657; PubMed=2474555;  
 RX Genbank G., Cibelli G., Rougon G., Mattei M.-G., Goridis C.;  
 RT "The mouse neuronal cell surface protein F3: a phosphatidylinositol-  
 anchored member of the immunoglobulin superfamily related to chicken  
 contactin.";  
 RT J. Cell Biol. 109:775-788 (1989).  
 RN [2]  
 RP MEDLINE=21289248; PubMed=11395001;  
 RX Boyle M.E., Berglund E.O., Mural K.K., Weber L., Peles E., Ranscht B.;  
 RT "Contactin orchestrates assembly of the septate-like junctions at the  
 paranode in myelinated peripheral nerve.";  
 RT Neuron 30:385-397 (2001).  
 CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS SYSTEM  
 DEVELOPMENT. IN ASSOCIATION WITH CTNNA1 SEEMS TO PLAY A ROLE IN  
 THE FORMATION OF PARANODAL AXO-GLIAL JUNCTIONS IN MYELINATED  
 PERIPHERAL NERVES AND MAY HAVE A ROLE IN THE SIGNALING BETWEEN  
 AXONS AND MYELINATING GLIAL CELLS.  
 CC -1- SUBUNIT: INTERACTS WITH CONTACTIN ASSOCIATED PROTEIN 1 IN CIS FORM  
 (BY SIMILARITY). Binds to the carbonic anhydrase like domain of  
 protein-tyrosine phosphatase zeta (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- MISCELLANEOUS: F3 SHARES WITH L1, N-CAM, MAG, AND OTHER CELL-  
 ADHESION MOLECULES FROM NERVOUS TISSUE THE L2/HNK-1 CARBOHYDRATE  
 EPITOPE.  
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 4 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).



FT LIPID 1001 1001 GPI-anchor amidated serine (Potential).  
SQ SEQUENCE 1021 AA; 113494 MW; FCDCL13055EE5C68 CRC64;  
Query Match 7.9%; Score 104; DB 1; Length 1021;  
Best Local Similarity 20.2%; Pred. No. 3;  
Matches 51; Conservative 28; Mismatches 89; Indels 84; Gaps 9;  
QY 16 LALLPAATQG--NKVVLGGKGVDELCTCTASQKKSIOFHWKNSNOIKILNQGSFLLTKGP 73  
DB 409 LALAPFFEMNPMKKKILAAAGKGVILECKPKAAPKPFSSV-----SKGT 452  
QY 74 SKLNDRAISRSLMDQGNFPLIKNKIEDSDTYICEVEQKGEVQLVGLTRANSPDTH 133  
DB 453 EMLVN--SSKILLWEDGS--LEINNIIRNDGCIYTCAENNRK-----ANS----- 495  
QY 134 LOGQSLLTLESPP-----GSSPSVQCRS----- 157  
DB 496 ---TGLVITNPFRIILAPINDITVGENATWQCAFPDPSLDLFFVMSFNGVIDFNK 551  
QY 158 -----PRGNIQGGKTLVSQLELQDSGTWCTVQONQKVEFKIDIVPRASALPAPP 210  
DB 552 EITNIHYQRNFMEDANGELLIRNAQLKHAGRYCTACTIVDSSASADLVVRGP--PGPP 609  
QY 211 TGSALPDPOTAS 222  
DB 610 GGLRIEDIRATS 621  
RESULT 55  
KVSS MOUSE  
ID KVSS MOUSE STANDARD; PRT; 108 AA.  
AC P01652;  
RT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region J606.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=82099361; PubMed=6798111;  
RA Johnson N., Slankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma inulin-  
binding proteins."  
RL J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT  
BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
DR PIR; A92811; KWS06.  
DR HSBP; P01607; IREI.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IgV\_1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23 FRAMEWORK-1.  
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 35 49 FRAMEWORK-2.  
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 57 88 FRAMEWORK-3.  
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 98 108 FRAMEWORK-4.  
FT DISULFID 23 88 BY SIMILARITY.  
FT NON TER 108  
SQ SEQUENCE 108 AA; 11810 MW; 8DE4DD31076F2AFB CRC64;  
Query Match 7.9%; Score 103.5; DB 1; Length 108;  
Best Local Similarity 39.0%; Pred. No. 0.2;  
Matches 30; Conservative 5; Mismatches 37; Indels 5; Gaps 2;  
QY 34 GDVTELTCTASQKKSIOFHWKNSNOIKILNQGSFLLTKGPSKLNDRADSRSLMDQG--NF 92

DB 16 GDIIVMTQASQGTSTINAMFQOKP-----GKAPKLLIYASNLDEGVSRFSGSRGTDF 71  
QY 93 PIIINKLKIEDSDTYIC 109  
DB 72 TLTISLEDEDMATYFC 88  
RESULT 56  
KVW HUMAN  
ID KVW HUMAN STANDARD; PRT; 129 AA.  
AC P04431;  
RT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region Walker precursor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85014148; PubMed=6091049;  
RA Klobeck H.G., Combrato G., Zachau H.G.;  
RT "Immunoglobulin genes of the kappa light chain type from two human  
lymphoid cell lines are closely related."  
RL Nucleic Acids Res. 12:6995-7006(1984).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by, and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL; X00965; CAA25477.1; ALT\_TERM.  
DR PIR; A01883; KIHMK.  
DR HSBP; P01607; IREI.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IgV\_1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 129 IG KAPPA CHAIN V-I REGION WALKER.  
FT DOMAIN 23 45 FRAMEWORK-1.  
FT DOMAIN 46 56 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 57 71 FRAMEWORK-2.  
FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 79 110 FRAMEWORK-3.  
FT DOMAIN 111 119 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 120 129 FRAMEWORK-4.  
FT DISULFID 45 110 BY SIMILARITY.  
FT NON TER 129  
SQ SEQUENCE 129 AA; 14069 MW; F941FA07DAFC2F9 CRC64;  
Query Match 7.9%; Score 103.5; DB 1; Length 129;  
Best Local Similarity 32.8%; Pred. No. 0.25; Indels 17; Gaps 5;  
Matches 43; Conservative 10; Mismatches 61;  
QY 1 MNRGVPFRHLILVQLALPAA-----TQGNKVVLGKGVDELCTCTASQKKSIOFHWK 54  
DB 1 MCKRVPAQ--LLGLLLMLRGARCDIQMTQSPESLSASVDRVTITCRASQSSINLVNWY 58  
QY 55 NSNQIKILNQGSFLLTKGPSKLNDRADSRSLMDQG--NFPLIKNKIEDSDTYICEVED 113  
DB 59 QOKP-----GKAPKLLIYASNLQSGVTSRFSGSGTDFLTLTSSLPQEDSATYFC----- 110



OY 114 QKEVQLLVFG 124  
 DB 111 QOSYSLTITFG 121  
 RESULT 57  
 CD7\_MOUSE STANDARD; PRT; 210 AA.  
 AC P50283;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell antigen CD7 precursor.  
 GN CD7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91138718; PubMed=7678579;  
 RA Yoshikawa K., Seto M., Ueda R., Odata Y., Fukatsu H., Segawa A.,  
 RT "Isolation and characterization of mouse CD7 cDNA."  
 RL Immunogenetics 37:114-119 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=94164701; PubMed=7509775;  
 RA Lee D.M., Watson M.L., Seldin M.F.;  
 RT "Mouse Cd7 maps to chromosome 11."  
 RL Immunogenetics 39:289-290 (1994).  
 RN [3]  
 RP SEQUENCE OF 1-24 AND 176-210 FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Spleen;  
 RX MEDLINE=95104926; PubMed=7528728;  
 RA Yoshikawa K., Seto M., Ueda R., Odata Y., Aoki S., Takahashi T.;  
 RT "Molecular cloning of the gene coding for the mouse T-cell antigen  
 CD7."  
 RL Immunogenetics 41:159-161 (1995).  
 RN [4]  
 RP INTERACTION WITH SECTM1.  
 RX MEDLINE=20119303; PubMed=10652336;  
 RA Lyman S.D., Escobar S., Rousseau A.-M., Armstrong A., Fanslow W.C.;  
 RT "Identification of CD7 as a cognate of the human K12 (SECTM1)  
 protein."  
 RL J. Biol. Chem. 275:3431-3437 (2000).  
 RT  
 CC -1- FUNCTION: Not yet known.  
 CC -1- SUBUNIT: Interacts with SECTM1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 CC  
 CC  
 DR EMBL; D10329; BAA01171.1; -  
 DR EMBL; U23462; BAA17482.1; -  
 DR EMBL; D31956; BAA06728.1; -  
 DR EMBL; D31957; BAA06728.1; JOINED.  
 DR EMBL; D31958; BAA06728.1; JOINED.  
 DR EMBL; D31959; BAA06728.1; JOINED.  
 DR PIR; I49294; I49294.  
 DR HSP; P80362; 1WTL.  
 DR MCD; MGI:88344; CD7.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003599; Ig.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00409; IG; 1.

DR PROSITE: P50835; IG LIKE; 1.  
 KW T-cell; Signal; Immune response; Antigen; Transmembrane; Glycoprotein;  
 KW Immunoglobulin domain; Receptor; Lipoprotein; Palmitate.  
 FT SIGNAL 1 23  
 FT CHAIN 24 210  
 FT DOMAIN 24 150  
 FT TRANSMEM 151 171  
 FT DOMAIN 172 210  
 FT DOMAIN 24 129  
 FT DISULFD 45 111  
 FT CARBOHYD 42 42  
 FT CARBOHYD 86 86  
 FT CARBOHYD 93 93  
 FT LIPID 168 168  
 FT CONFLICT 69 69  
 SQ SEQUENCE 210 AA; 23152 MW; 0EBD968B302962B CRC64;  
 Query Match 7.9%; Score 103.5; DB 1; Length 210;  
 Best Local Similarity 21.7%; Pred. No. 0.46;  
 Matches 57; Conservative 32; Mismatches 61; Indels 113; Gaps 13;  
 OY 10 LLLVQLA-LIPA-----ATQGNKVVLGKGDVVELCTASQKKSITQFMKNSNQIKIL 62  
 DB 7 LALLTLTAGILGPELPDQDVHQSPLTASGDSVNTICSTR----- 48  
 OY 63 GNGSGFLTKGSPKLNDRADSRSLMDQGNFLLIKNLIKEDSDTYICEVEDQKEVQLLV 122  
 DB 49 GHLGELIMK-----KIWPA-----YNYVFEKRE----- 74  
 OY 123 FGLTRANSDTHLQGSULTLESPGSSPSVQCRSPRGK-NIQGK---TLVSQLELD 178  
 DB 75 -----PTVD-RTFSGRINFSSQKNLTTISLQAD 105  
 OY 179 SGWTCTVLQNKVY-----FKDIVPRAALPAPPGSALPDQTSALPDPPAAS-AL 233  
 DB 106 TGDYTCBAV---RKVSARGLFTTVVKEKSS-----QEAYRSQEPLOTSPSF 149  
 OY 234 PALAVISFLGLGIVACVILAR 256  
 DB 150 PAATAVGFFFTGLLGVVCSMLR 172  
 RESULT 58  
 NRG DROME  
 ID\_NRG DROME STANDARD; PRT; 1302 AA.  
 AC P20241; O61541; O61542; Q24414; Q24415; Q95064; Q9V3X0;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neuroglial precursor.  
 GN NRG OR CG1634.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND SEQUENCE OF 24-41 AND 737-751.  
 RX MEDLINE=90030418; PubMed=2805067;  
 RA Bieber A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,  
 RA Triquina Z.R., Schilling J., Goodman C.S.;  
 RT "Drosophila neuroglial: a member of the immunoglobulin superfamily  
 RT with extensive homology to the vertebrate neural adhesion molecule  
 RT L1."  
 RL Cell 59:447-460 (1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE=98332718; PubMed=9666073;  
 RA Zhao G., Hortsch M.;  
 RT "The analysis of genomic structures in the L1 family of cell adhesion  
 RT proteins provides no evidence for exon shuffling events after the  
 RT separation of arthropod and chordate lineages."  
 RL Gene 215:47-55 (1998).

[31]  
RP REVISIONS.  
RA Horteck M.,  
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.

[4]  
RN SEQUENCE FROM N.A. (ISOFORM SHORT).  
RP STRAIN=Berkley;  
RC MEDLINE=20196006; PubMed=10731132;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RZ Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Baller R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brockstein F., Brotler P.,  
RA Butte K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Paolis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heilmann T.J., Hernandez J.R., Houck J.,  
RA Howland D., Houston K.A., Howland T.J., Wei M.-H., Ibgwen C.,  
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Koditz C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Mosher A.I.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirydas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassenaar D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yah R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.",  
RL Science 287:2185-2195(2000).

[5]  
RN SEQUENCE FROM N.A. (ISOFORM SHORT).  
RP STRAIN=Berkley; TISSUE=Head.  
RC STAPLETON M., BROKSTEIN P., HONG L., AGBAYANI A., CARLSON J.,  
RA CHAMPE M., CHAVEZ C., DORSETT V., FARLAN D., FRISSE E., GEORGE R.,  
RA GONZALEZ M., GUARIN H., LI P., LIAO G., MIRANDA A., MUNGILL C.J.,  
RA NUOVO J., PACLE J., PARAGAS V., PARK S., PHOUENAVONG S., WAN K.,  
RA YUO C., LEWIS S.E., RUBIN G.M., CELNIKER S.,  
RL submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

[6]  
RN SEQUENCE OF 1182-1302 FROM N.A., FUNCTION, ALTERNATIVE SPLICING, AND  
RP TISSUE SPECIFICITY.  
RC TISSUE=Embryo;  
RX MEDLINE=90268720; PubMed=1693086;  
RA Horteck M., Bieber A.J., Patel N.H., Goodman C.S.;  
RT "Differential splicing generates a nervous system-specific form of  
RL *Drosophila neuroglian*.",  
RL Neuron 4:697-709(1990).

[7]  
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.  
RP MEDLINE=94231741; PubMed=1512815;  
RA Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;  
RT "Crystal structure of tandem type III fibronectin domains from  
RL *Drosophila neuroglian* at 2.0 Å.",  
RL Neuron 12:717-731(1994).

-I-  
FUNCTION: The long isoform may play a role in neural and glial  
cell adhesion in the developing embryo. The short isoform may be a

[illegible]

FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1224 1239 OPTEDSGSTIGQVPEK -> MNEDSGSTIGQVPEK (in isoform short).  
 FT VARSPLIC 1240 1302 /FTID=VSP\_002601.  
 FT VARSPLIC 1240 1302 Missing (in isoform short).  
 FT CONFLICT 85 86 NR -> KP (in REF. 2).  
 FT CONFLICT 1282 1282 /FTID=VSP\_002602.  
 FT STRAND 619 625 MISSING (in REF. 6).  
 FT STRAND 629 635  
 FT TURN 640 641  
 FT STRAND 646 653  
 FT TURN 657 658  
 FT STRAND 661 668  
 FT TURN 669 670  
 FT STRAND 673 677  
 FT STRAND 682 692  
 FT TURN 693 694  
 FT STRAND 695 696  
 FT STRAND 706 708  
 FT STRAND 721 723  
 FT TURN 727 728

Query Match 7.9%, Score 103.5; DB 1; Length 1302;  
 Best Local Similarity 20.7%, Pred. No. 4.3; Matches 54; Conservative 38; Mismatches 104; Indels 65; Gaps 13;

QY 25 GNNKVLGKKGDPVLELTCTASQKK--SIQFMKNSNQIKILNQGS-FLTKGSPSKLNDRA- 80  
 DB 225 GNNKVL-----DYKMGVSAQNKHPVQVYRSRQSLALRKRMELFPIYGTPLPQVW 280  
 QY 81 --DSRRSLMD---QGNF--PLIKNLIKEDSDTYICEVEDQKEVQLVFGLTANSDFH 132  
 DB 281 SKDCQRIQMSDRITQGHYKSLVIRQTFDDAGTYTCVNSGVNAQSFIIINWNSVPY 340  
 QY 133 LLQGSULTLESPPGSSPSVQCRS-----FRGNKIQGK-----TISV 171  
 DB 341 FTKEPELITAAED--EEVFECPRAAGVPEPKISWINGKPEIESTPNRRTVTDNTRI 397  
 QY 172 SQLELDSDGTCTCTVLQNKQKVEFKIDIVPRASALPDPGTASALPDPAPAS 231  
 DB 398 INLVKGDVGNVCNA-TNSLGVYK-DVYLVNQ-----EPPIS 435

QY 232 ALPALAVISFLGLGLGVAC 252  
 DB 436 EAPAAVSTVD---GRNVTKIC 453

RESULT 59  
 ECTO\_RAT STANDARD; PRT; 519 AA.  
 AC P16573;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ecto-ATPase precursor (Cell-CAM 105) (C-CAM 105) (ATP-dependent  
 DE taurocholate-carrier protein) (GP110).  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 50-68.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=893140561; PubMed=2527235;  
 RA Lin S.-H., Guidotti G.,  
 RT "Cloning and expression of a cDNA coding for a rat liver plasma  
 RT membrane ecto-ATPase. The primary structure of the ecto-ATPase is  
 RT similar to that of the human biliary glycoprotein I.",  
 RL J. Biol. Chem. 264:14408-14414 (1989).

RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=Sprague-Dawley, and Wistar; TISSUE=Liver;  
 RX MEDLINE=93279310; PubMed=8504806;  
 RA Edlund M., Gaardsvoll H., Bock E., Oebirink B.,  
 RT "Different isoforms and stock-specific variants of the cell adhesion  
 RT molecule C-CAM (cell-CAM 105) in rat liver.",  
 RL Eur. J. Biochem. 213:1109-1116 (1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=92344597; PubMed=1637321;  
 RA Culic O., Huang Q., Flanagan D., Hixson D., Lin S.-H.,  
 RT "Molecular cloning and expression of a new rat liver cell-CAM105  
 RT isoform. Differential phosphorylation of isoforms.",  
 RL Biochem. J. 285:47-53 (1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=94058980; PubMed=8240240;  
 RA Cheung P.H., Culic O., Qiu Y., Earley K., Thompson N., Hixson D.C.,  
 RT Lin S.-H.,  
 RT "The cytoplasmic domain of C-CAM is required for C-CAM-mediated  
 RT adhesion function: studies of a C-CAM transcript containing an  
 RT unspliced intron.",  
 RL Biochem. J. 295:427-435 (1993).  
 RN [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=9029222; PubMed=2141577;  
 RA Arivillius M., Hansen O.C., Lazrek M.B.S., Bock E., Oebirink B.,  
 RT "The cell adhesion molecule Cell-CAM 105 is an ecto-ATPase and a  
 RT member of the immunoglobulin superfamily.",  
 RL FEBS Lett. 264:267-269 (1990).  
 RN [6]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=Wistar; TISSUE=Liver;  
 RX MEDLINE=93292517; PubMed=8513803;  
 RA Becker A., Lucka L., Kilian C., Kamnitsch C., Reuter W.,  
 RT "Characterisation of the ATP-dependent taurocholate-carrier protein  
 RT (GP110) of the hepatocyte canalicular membrane.",  
 RL Eur. J. Biochem. 214:539-548 (1993).  
 RN [7]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=93123234; PubMed=8380406;  
 RA Najjar S.M., Accilli D., Philippe N., Jernberg J., Margolis R.,  
 RA Taylor S.I.,  
 RT "p120/ecto-ATPase, an endogenous substrate of the insulin receptor  
 RT tyrosine kinase, is expressed as two variably spliced isoforms.",  
 RL J. Biol. Chem. 268:1201-1206 (1993).  
 RN [8]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91354197; PubMed=1831973;  
 RA Lin S.-H., Culic O., Flanagan D., Hixson D.C.,  
 RT "Immunohistochemical characterization of two isoforms of rat liver ecto-  
 RT ATPase that show an immunological and structural identity with a  
 RT glycoprotein cell-adhesion molecule with Mr 105,000.",  
 RL Biochem. J. 278:155-161 (1991).  
 CC -1 FUNCTION: THIS PROTEIN IS A CALCIUM-INDEPENDENT CELL ADHESION  
 CC MOLECULE WITH HOMOPHILIC BINDING PROPERTIES. MAY PLAY A ROLE IN  
 CC THE FORMATION AND MAINTENANCE OF THE SPECIALIZED MEMBRANE  
 CC STRUCTURE OF THE APICAL SURFACE OF THE HEPATOCYTES.  
 CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CANALICULAR DOMAIN  
 CC OF HEPATOCYTE PLASMA MEMBRANS.  
 CC -1 ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named Isoforms=2;  
 CC Name=Long;  
 CC IsoId=P16573-1; Sequence=Displayed;  
 CC Name=Short; Synonyms=C-CAM3;  
 CC IsoId=P16573-2; Sequence=VSP\_002504, VSP\_002505;  
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN EPITHELIA, VESSEL ENDOTHELIA,  
 CC LEUKOCYTES AND PLATELETS.  
 CC -1 PPM: PHOSPHORYLATED ON SERINE RESIDUES IN BOTH ISOFORMS. ON SERINE  
 CC AND TYROSINE RESIDUES IN CYTOPLASMIC REGION OF THE LONG ISOFORM.

CC	-	SIMILARITY:	Belongs to the immunoglobulin superfamily. CEA family.
CC	-	SIMILARITY:	Contains 3 immunoglobulin-like C2-type domains.
CC	-	SIMILARITY:	Contains 1 immunoglobulin-like V-type domain.
CC	-	This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).	
DR	EMBL; J04963;	AAA41104.1; .-	
DR	EMBL; X71122;	CAN50435.1; .-	
DR	EMBL; Z12019;	CAA78054.1; .-	
DR	EMBL; M92848;	AAAI6783.1; ALT_TERM.	
DR	PIR; A44783;	A44783.	
DR	PIR; S23969;	S23969.	
DR	InterPro: IPR007110;	Ig-Like.	
DR	InterPro: IPR003598;	Ig_c2.	
DR	Pfam; PF00047;	Ig_3.	
DR	SMART; SM00408;	IGc2; 2.	
DR	PROSITE; PS50835;	IG LIKE; 3.	
KW	Cell adhesion; Transmembrane; Glycoprotein; Immunoglobulin domain;		
KV	Repeat; Signal; Phosphorylation; Alternative splicing;		
KM	Pyroglutamate carboxylic acid.		
FT	SIGNAL	1..34	POTENTIAL.
FT	CHAIN	35..519	Ecto-ATPase.
FT	DOMAIN	35..422	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	423..448	POTENTIAL.
FT	DOMAIN	449..519	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	42..140	IG-LIKE V-TYPE.
FT	DOMAIN	147..232	IG-LIKE C2-TYPE 1.
FT	DOMAIN	237..317	IG-LIKE C2-TYPE 2.
FT	DOMAIN	325..403	IG-LIKE C2-TYPE 3.
FT	DISULFID	167..215	BY SIMILARITY.
FT	DISULFID	259..299	BY SIMILARITY.
FT	DISULFID	344..392	BY SIMILARITY.
FT	MOD_RES	35..35	PYROGLUTAMATE CARBOXYLIC ACID
FT	CARBOHYD	87..87	(N-LINKED GLCNAC...)
FT	CARBOHYD	104..104	(N-LINKED GLCNAC...)
FT	CARBOHYD	113..113	(N-LINKED GLCNAC...)
FT	CARBOHYD	148..148	(N-LINKED GLCNAC...)
FT	CARBOHYD	152..152	(N-LINKED GLCNAC...)
FT	CARBOHYD	172..172	(N-LINKED GLCNAC...)
FT	CARBOHYD	173..173	(N-LINKED GLCNAC...)
FT	CARBOHYD	197..197	(N-LINKED GLCNAC...)
FT	CARBOHYD	224..224	(N-LINKED GLCNAC...)
FT	CARBOHYD	256..256	(N-LINKED GLCNAC...)
FT	CARBOHYD	288..288	(N-LINKED GLCNAC...)
FT	CARBOHYD	292..292	(N-LINKED GLCNAC...)
FT	CARBOHYD	302..302	(N-LINKED GLCNAC...)
FT	CARBOHYD	315..315	(N-LINKED GLCNAC...)
FT	CARBOHYD	331..331	(N-LINKED GLCNAC...)
FT	CARBOHYD	374..374	(N-LINKED GLCNAC...)
FT	VASPPPLIC	455..458	GSDH -> GSP (in isoform Short).
FT	VASPPPLIC	455..458	/FTId=VSP_002504.
FT	VARSPLIC	459..519	Mising (in isoform Short).
FT	VARSPLIC	459..519	/FTId=Vsp_002505.
FT	VARIANT	49..49	S->K.
FT	VARIANT	55..55	T->A.
FT	VARIANT	70..70	V->G.
FT	VARIANT	73..76	TGLIN -> LNPd.
FT	VARIANT	86..86	S->D.
FT	VARIANT	88..88	T->M.
FT	VARIANT	90..90	Q->K.
FT	VARIANT	92..92	E->G.
FT	VARIANT	99..99	V->E.
FT	VARIANT	118..119	GP -> RA.
FT	VARIANT	125..125	I->F.
FT	VARIANT	127..127	K->Q.
FO	SEQUENCE	519 AA; 57263 MM; 041439280D7975776 CRC64;	

```

Query March 7.8%; Score 103; DB 1; Length 519;
Best Local Similarity 20.9%; Prod. No. 1.5;
Matches 51; Conservative 34; Mismatches 93; Indels 66; Gaps 7;

Oy 33 KGDIVLTCTASOKKSIQIFHWKNSNOIKILGNQGSFLTGPSPKLNDRADSRRLMDQGNF 92
Db 251 QGSNLNTSCHADSNPAPQYFWMLINEKLTQSSQE-----SDPHLL 134
Oy 93 PLIKNLKIDSDPTTICEVDQKEEYQVLVFGILAN-----SDPHLL 134
Db 284 -LFTSNITNTNNGSYACFVANN-----TVTGSRTTVKNITVEEPTQPSIQITNTTVK 335
Oy 135 QGQSLVLTLESPPG-----SSPSVQCRSPRGKNIQGQKTLVSQLEFDSDGTWCTV 186
Db 336 ELGSLVLTCTCSKOTGVSVKRLFNQSGLQ-TDNTMLTSQDMSITLIPDIKREDADYCEI 394
Oy 187 LQNKQKVEFKIDIVPRASALPAPPTG-SALPDQTSALPDPPAPASLPALAVISFLG 245
Db 395 ---SNPFSFRISHPIKDIVPDPQNGSGSEGLAGIVIGSAGVALPALAVFYSRK 451
Oy 246 LGLG 249
Db 452 TGGG 455

RESULT 60
LAG3_MOUSE
ID_LAG3_MOUSE STANDARD; PRT; 521 AA.
AC Q61790.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DR Lymphocyte activation gene-3 protein precursor (LAG-3).
GN LAG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RC MEDLINE=97079438; PubMed=8921170.
RA Mestirengeli R., Micanelli E., Donini S.;
RT "Cloning of murine LAG-3 by magnetic bead bound homologous probes and
RT PCR (gene-capture PCR).";
RL Anal. Biochem. 241:93-102(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RC MEDLINE=96185498; PubMed=8602528;
RA Miyazaki T., Dierich A., Benoist C., Mathis D.;
RT "Independent modes of natural killing distinguished in mice lacking
RT Lag3.";
RL Science 272:405-408(1996).
CC -1- FUNCTION: Involved in lymphocyte activation. Binds to HLA class-II
CC antigens (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X98113; CAA66794.1; -.
DR MGD: MGI:106588; Lag3.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003599; Ig.
DR

```

DR Pfam: PF00047; Ig; 2.  
 DR SMART: SMO0409; IG; 3.  
 DR PROSITE: PS00835; IG-LIKE; 2.  
 KM Signal1, Transmembrane; Glycoprotein; Immunoglobulin domain; Repeat.  
 FT SIGNAL 1 22  
 FT CHAIN 23 521  
 FT DOMAIN 23 442  
 FT TRANSMEM 443 463  
 FT DOMAIN 464 521  
 FT DOMAIN 37 163  
 FT DOMAIN 165 246  
 FT DOMAIN 258 341  
 FT DOMAIN 345 412  
 FT DOMAIN 505 518  
 FT DISULFID 44 156  
 FT DISULFID 185 235  
 FT DISULFID 276 327  
 FT DISULFID 363 405  
 FT CARBOHYD 184 184  
 FT CARBOHYD 244 244  
 FT CARBOHYD 309 309  
 FT CARBOHYD 337 337  
 FT CARBOHYD 381 381  
 SQ SEQUENCE 521 AA; 56977 MM; 64C29C40EB780E48 CRC64;  
 Query March 7.8%; Score 103; DB 1; Length 521;  
 Best Local Similarity 24.9%; Pred. No. 1.5; Mismatches 96; Indels 32; Gaps 11;  
 Matches 55; Conservative 38; Mismatches 96; Indels 32; Gaps 11;  
 QY 5 VPFRHLVLQIALPAATQGNKVLGKGGDTVELTCTASQ-KKSIGPFH-KNSNQIKIL 62  
 DB 149 LFNALSGSLRLRGQASMINSPSGVLKLSQWVLNCSFSPDRPVSHMGQGNRPVY 208  
 QY 63 GNQGSFLTKGSKLNDRADSRSLMDQGNFPLIK-----NLKEDSTYICEVEDQKEE 117  
 DB 209 NSPHFLAETFLLL-----POVSPIDSGTCGCVLYRDGFVSI-----TYNLKVLGLEPV 259  
 QY 118 VQLVFE---GLTANSDTLLQSG---QSLTLTLESPPGSSPVQCRSPRKGIOGKTLVS 171  
 DB 260 APLTVVAEGSRVLPCHLPVGVGPSLLINKMPVGGGPELPV--AGKS--GNFTLHL 314  
 QY 172 SQLELDGSGTWTCTV-LQNRK---KYEFKIDIVPRASALP 207  
 DB 315 EAVGLAAGCTYTCGSHLQGGQNLMTAVLTVPKSGLP 355  
 RESULT 61  
 ID CD22 MOUSE STANDARD: PRT: 862 AA.  
 AC P31329; O9JUH2; O9JUX9; O9JUY0; O9JUY1; O9R056; O9R094; O9WU51;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE B-cell receptor CD22 precursor (Ieu-14) (B-lymphocyte cell adhesion molecule) (BR-CAM) (Siglec-2).  
 GN CD22 OR LYB-8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RX NCBI\_TaxId=10090;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DMB/2J, and BALB/c; TISSUE=Liver;  
 RX MEDLINE=93315834; PubMed=8100843;  
 RA Brenan C.I., Copeland N.G., Jenkins N.A., Clack E.A.;  
 RT "Organization of the murine Cd22 locus. Mapping to chromosome 7 and characterization of two alleles."  
 RL J. Immunol. 151:175-187(1993).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6, BxSB, and MRL;  
 RX MEDLINE=99432012; PubMed=10501843;

RA Lajthaia F., Ikon-Zekri N., Fossati Jimack L., Chicheportiche Y.,  
 RA Parkhouse R.M., Mary C., Reininger L., Brigson G., Izui S.;  
 RT "Polymorphisms in the Cd22 gene of inbred mouse strains."  
 RL Immunogenetics 49:991-995(1999).  
 RN (3)  
 RP SEQUENCE OF 1-178 FROM N.A. (ISOFORMS 1; 2 AND 3).  
 RC STRAIN=NZW; TISSUE=Spleen;  
 RX MEDLINE=20432275; PubMed=10975807;  
 RA Mary C., Laporte C., Parzy D., Santiago M.L., Stefani F.,  
 RA Lajthaia F., Parkhouse M.E., O'Keefe T.L., Neuberger M.S., Izui S.,  
 RA Reininger L.;  
 RT "Dysregulated expression of the Cd22 gene as a result of a short interspersed nucleotide element insertion in Cd22alpha lupus-prone mice."  
 RL J. Immunol. 165:2987-2996(2000).  
 RN (4)  
 RP SIALIC ACID BINDING.  
 RX MEDLINE=95179521; PubMed=7533044;  
 RA Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,  
 RA Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;  
 RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily."  
 RL Curr. Biol. 4:965-972(1994).  
 RN (5)  
 RP INTERACTION WITH GRR2, SYK, PIK3R1/PIK3R2 AND PLCG1, PHOSPHORYLATION OF TYR-777, TYR-822, TYR-837 AND TYR-857, AND MUTAGENESIS OF TYR-822.  
 RX MEDLINE=99303650; PubMed=10373493;  
 RA Yohannan J., Wienands J., Coggeshall K.M., Justement L.B.;  
 RT "Analysis of tyrosine phosphorylation-dependent interactions between stimulatory effector proteins and the B cell co-receptor CD22."  
 RL J. Biol. Chem. 274:18769-18776(1999).  
 RN (6)  
 RP INTERACTION WITH GRR2, SHC1 AND INPP5D, AND PHOSPHORYLATION BY LYN.  
 RX MEDLINE=20298808; PubMed=10748054;  
 RA Poe J.C., Fujimoto M., Jansen P.J., Miller A.S., Tedder T.F.;  
 RT "CD22 forms a quaternary complex with SHP, Gb2, and Shc. A pathway for regulation of B lymphocyte antigen receptor-induced calcium flux."  
 RL J. Biol. Chem. 275:17420-17427(2000).  
 CC -1- FUNCTION: Mediates B-cell B-cell interactions. May be involved in the localization of B-cells in lymphoid tissues. Binds sialylated glycoproteins; one of which is CD45. Preferentially binds to alpha2,6-linked sialic acid. The sialic acid recognition site can be masked by cis interactions with sialic acids on the same cell surface. Upon ligand induced tyrosine phosphorylation in the immune response seems to be involved in regulation of B cell antigen receptor signaling. Plays a role in positive regulation through interaction with Src family tyrosine kinases and may also act as an inhibitory receptor by recruiting cytoplasmic phosphatases via their SH2 domains that block signal transduction through dephosphorylation of signaling molecules.  
 CC -1- SUBUNIT: Interacts with SYK, PIK3R1/PIK3R2, PLCG1, SHC1, INPP5D and GRR2 upon phosphorylation. May form a complex with INPP5D/SHP, GRR2 and SHC1. Interacts with PTPN6/SHP-1 upon phosphorylation (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1; Synonym=CD22-beta;  
 CC IsoId=P35329-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P35329-2; Sequence=VSP\_002532;  
 CC Name=3;  
 CC IsoId=P35329-3; Sequence=VSP\_002533;  
 CC -1- TISSUE SPECIFICITY: B lymphocytes.  
 CC -1- DOMAIN: Contains 3 copies of a cytoplasmic motif that is referred to as the immunoreceptor tyrosine-based inhibitor motif (ITIM). This motif is involved in downmodulation of cellular responses. The phosphorylated ITIM motif binds to the SH2 domain of PTPN6/SHP-1.  
 CC -1- PTM: Phosphorylated on tyrosine residues by LYN (Probable).  
 CC -1- PTM: Phosphorylation on Tyr-822 is involved in the binding to

Query Match	Best Local Similarity	7.8%;	Score 102.5;	DB 1;	Length 862;
Matches	52;	Conservative	21.9%;	Pred. No. 3.1;	Mismatches 85;
					Indels 67;
					Gaps 10;
FT CARBOHYD	378	378	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	408	408	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	460	460	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	561	561	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	589	589	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT VASPLIC	54	115	Missing (in isoform 2).		
FT VASPLIC	100	133	/Frid=VSP_002532.		
FT VASPLIC	15	15	Missing (in isoform 3).		
FT VASPLIC	15	15	/Frid=VSP_002533.		
FT VASPLIC	15	15	A -> V (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	19	19	Q -> R (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	76	76	K -> T (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	83	86	KAP -> ATRKEDPS (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	90	91	PP -> LS (IN BALB/C).		
FT VASPLIC	94	94	R -> G (IN BALB/C).		
FT VASPLIC	102	102	S -> R (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	112	114	PIR -> LII (IN BXS).		
FT VASPLIC	173	173	E -> G (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	179	179	Q -> K (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	186	186	K -> E (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	190	190	V -> I (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	192	192	P -> S (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	236	236	E -> K (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	241	241	R -> C (IN BALB/C).		
FT VASPLIC	244	244	R -> H (IN BALB/C AND MRL).		
Query Match	Best Local Similarity	7.8%;	Score 102.5;	DB 1;	Length 862;
Matches	52;	Conservative	21.9%;	Pred. No. 3.1;	Mismatches 85;
					Indels 67;
					Gaps 10;
FT CARBOHYD	378	378	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	408	408	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	460	460	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	561	561	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	589	589	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT VASPLIC	54	115	Missing (in isoform 2).		
FT VASPLIC	100	133	/Frid=VSP_002532.		
FT VASPLIC	15	15	Missing (in isoform 3).		
FT VASPLIC	15	15	/Frid=VSP_002533.		
FT VASPLIC	15	15	A -> V (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	19	19	Q -> R (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	76	76	K -> T (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	83	86	KAP -> ATRKEDPS (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	90	91	PP -> LS (IN BALB/C).		
FT VASPLIC	94	94	R -> G (IN BALB/C).		
FT VASPLIC	102	102	S -> R (IN BALB/C, C57BL/6, BXS AND MRL).		
FT VASPLIC	112	114	PIR -> LII (IN BXS).		
FT VASPLIC	173	173	E -> G (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	179	179	Q -> K (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	186	186	K -> E (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	190	190	V -> I (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	192	192	P -> S (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	236	236	E -> K (IN BALB/C, C57BL/6 AND MRL).		
FT VASPLIC	241	241	R -> C (IN BALB/C).		
FT VASPLIC	244	244	R -> H (IN BALB/C AND MRL).		
Query Match	Best Local Similarity	7.8%;	Score 102.5;	DB 1;	Length 862;
Matches	52;	Conservative	21.9%;	Pred. No. 3.1;	Mismatches 85;
					Indels 67;
					Gaps 10;
FT CARBOHYD	378	378	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	408	408	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	460	460	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	561	561	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	589	589	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT VASPLIC	54	115	Missing (in isoform		

RL Gene 174:3-8(1996).  
 RP [2]  
 RC SEQUENCE OF 910-1348 FROM N.A.  
 RA TISSUE-Spinal cord;  
 RX MEDLINE=93378866; PubMed=8396413;  
 RA Elchmann A., Marcelle C., Breant C., Le Douarin N.M.;  
 RT "Two molecules related to the VEGF receptor are expressed in early  
 RL endothelial cells during avian embryonic development."; Mech. Dev. 42:33-48(1993).  
 RN [3]  
 RP SEQUENCE OF 764-880 FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE-Embryo;  
 RX MEDLINE=95301109; PubMed=7781909;  
 RA Flamme I., Breier G., Risau W.;  
 RT "Vascular endothelial growth factor (VEGF) and VEGF receptor 2 (Flk-1)  
 RL are expressed during vasculogenesis and vascular differentiation in  
 RT the quail embryo."; Dev. Biol. 169:699-712(1995).  
 CC -1- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN  
 CC KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM  
 CC PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF  
 CC VASCULAR PERMEABILITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: IN ALL ENDOTHELIAL TISSUES DURING ONSET OF  
 CC VASCULARIZATION. IN LATER DEVELOPMENT, PRESENT IN LUNG, HEART,  
 CC INTESTINE AND SKIN.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE MESODERM AT ONSET OF  
 CC GASTRULATION. FROM DAY 2, CONFINED TO ENDOTHELIAL TISSUES AND  
 CC EXPRESSION CONTINUES TO BE WIDESPREAD THROUGHOUT VASCULARIZATION  
 CC UNTIL E9 WHERE IT BECOMES RESTRICTED TO SPECIFIC REGIONS SUCH AS  
 CC THE SIGNAL CHORD AND HEART VALVES.  
 CC -1- INDUCTION: IN VITRO, VEGF IS INDUCED BY BASIC FIBROBLAST GROWTH  
 CC FACTOR (bFGF), UNIQUELY IN THE FIRST 24 H OF CELL CULTURE.  
 CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.  
 CC CSF-1/PDGF receptor subfamily.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; X83288; CA58268.1; -;  
 DR EMBL; S65205; AAB28127.1; -;  
 DR EMBL; S78345; AAB34594.1; -;  
 DR PIR; JC4953; S51656.  
 DR HSP; P11362; IFGK.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR001824; RecepttyrkineIII.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR008266; Tyr\_kinase\_AS.  
 DR Pfam; PF00047; Ig; 6.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 2.  
 DR SMART; SM00408; IGC2; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00835; IG-LIKE; 5.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 DR Angiogenesis; Signal; transferase; tyrosine-protein kinase; Receptor;  
 KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 KW Immunoglobulin domain; Repeat.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 1348 VASCULAR ENDOTHELIAL GROWTH FACTOR

FT FT DOMAIN 21 756 RECEPTOR 2.  
 FT FT TRANSMEM 757 777 EXTRACELLULAR (POTENTIAL).  
 FT FT DOMAIN 778 1348 POTENTIAL.  
 FT FT DOMAIN 43 106 CYTOPLASMIC (POTENTIAL).  
 FT FT DOMAIN 138 202 IG-LIKE C2-TYPE 1.  
 FT FT DOMAIN 220 312 IG-LIKE C2-TYPE 2.  
 FT FT DOMAIN 320 405 IG-LIKE C2-TYPE 3.  
 FT FT DOMAIN 412 534 IG-LIKE C2-TYPE 4.  
 FT FT DOMAIN 540 651 IG-LIKE C2-TYPE 5.  
 FT FT DOMAIN 658 744 IG-LIKE C2-TYPE 6.  
 FT FT DOMAIN 825 1155 IG-LIKE C2-TYPE 7.  
 FT FT NP BIND 831 839 PROTEIN KINASE.  
 FT FT BINDING 859 859 ATP (BY SIMILARITY).  
 FT FT ACT SITE 1021 1021 ATP (BY SIMILARITY).  
 FT FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT FT CONFLICT 865 865 A -> S (IN REF. 3).  
 SQ SEQUENCE 1348 AA; 150305 MW; ASE4194A76FDSFB3 CRC64;  
 Query Match 7.88; Score 102.5; DB 1; Length 1348;  
 Best Local Similarity 22.98; Pred. No. 5.4;  
 Matches 49; Conservative 33; Mismatches 73; Indels 59; Gaps 12;  
 QY 9 HLLVLIQALLPATQGNKVVGLKGGTVLTCASQKSIQTHW-KNSNQIKLNGQS 67  
 DB 536 HVTRGLINIQPSQLTEK-----DNISLQCTADTFTEKLSWYLSLTHV--SQTP 584  
 QY 68 FLTKG--PSKINDRAUSRRSL-----WDQGNFPIILNKLKIEDSTYICEVEDQKEE 117  
 DB 585 F---GGIPMPVCNKLDAQLQKNTATVSNVNGENTLLELNISLQDGGDVLCAQDKKAK 641  
 QY 118 VQ-LVYFGLTANSDFH---LLQGSLLTLESPPGSSPSVOCRSFRG-----KN 162  
 DB 642 TORCLVHGLTVQEPPLHRLVGNLENQTTNI-----GETTEVLC-TVNGVPPNITWFKN 694  
 QY 163 IQ-----GGKTLVSQLEHLDGSTWTC 184  
 DB 695 SETLFEDSGIVLKDGNKTLTIRVRKEDGGLYTC 728  
 RESULT 63  
 CEAL\_MOUSE STANDARD; PRT; 521 AA.  
 ID CEAL\_MOUSE  
 AC P31809;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Carcinoembryonic antigen-related cell adhesion molecule 1 precursor  
 DE (Biliary glycoprotein 1) (BGP-1) (Murine hepatitis virus receptor)  
 DE (MHV-R) (Biliary glycoprotein D).  
 GN CEACAM1 OR BGP OR BGP1 OR BGPD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Colon;  
 RX MEDLINE=93100785; PubMed=9380065;  
 RA Dvorkin G.S., Dieffenbach C.B., Cardelliichio C.B., McCuaig K.,  
 RA Pensiero M.N.,  
 RT "Several members of the mouse carcinoembryonic antigen-related  
 RT glycoprotein family are functional receptors for the coronavirus  
 RT mouse hepatitis virus-A59.";  
 RL J. Virol. 67:1-8(1993).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=92046352; PubMed=1719235;  
 RA Dvorkin G.S., Pensiero M.N., Cardelliichio C.B., Williams R.K.,  
 RA Jiang G.-S., Holmes K.V., Dieffenbach C.W.,  
 RT "Cloning of the mouse hepatitis virus (MHV) receptor: expression in  
 RT human and hamster cell lines confers susceptibility to MHV.";  
 RL J. Virol. 65:6881-6891(1991).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93273228; PubMed=8500759;  
 RA McCuaig K., Rosenberg M., Nedellec P., Turbide C., Beauchemin N.,  
 RT "Expression of the Bgp gene and characterization of mouse colon  
 RT biliary glycoprotein isoforms.";  
 RL Gene 127:173-183(1993).  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=CD-1; TISSUE=Colon;  
 RX MEDLINE=89195121; PubMed=2702644;  
 RA Beauchemin N., Turbide C., Afar D., Raymond M., Bell J.,  
 RA Stammers C.P., Fuks A.,  
 RT "A mouse analogue of the human carcinoembryonic antigen.";  
 RL Cancer Res. 49:2017-2021(1989).  
 RN (5)  
 RP SEQUENCE OF 35-59.  
 RX MEDLINE=91288498; PubMed=1648219;  
 RA Williams R.K., Jiang G.-S., Holmes K.V.,  
 RT "Receptor for mouse hepatitis virus is a member of the  
 RT carcinoembryonic antigen family of glycoproteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5533-5536(1991).  
 CC -1- FUNCTION: Unknown: receptor for murine coronavirus MHV-A59.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P31809-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P31809-2; Sequence=VSP\_002484, VSP\_002485;  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

FT SIGNAL 1 34  
 FT CHAIN 35 521  
 FT DOMAIN 35 428  
 FT TRANSMEM 429 447  
 FT DOMAIN 448 521  
 FT DOMAIN 35 142  
 FT DOMAIN 147 234  
 FT DOMAIN 239 319  
 FT DOMAIN 323 411  
 FT DISULFID 167 217  
 FT DISULFID 261 301  
 FT DISULFID 346 394  
 FT CARBOHYD 71 71  
 FT CARBOHYD 89 89  
 FT CARBOHYD 104 104  
 FT CARBOHYD 148 148  
 FT CARBOHYD 152 152  
 FT CARBOHYD 199 199  
 FT CARBOHYD 206 206  
 FT CARBOHYD 210 210  
 FT CARBOHYD 226 226  
 FT CARBOHYD 258 258  
 FT CARBOHYD 290 290  
 FT CARBOHYD 294 294  
 FT CARBOHYD 304 304  
 FT CARBOHYD 317 317  
 FT CARBOHYD 333 333  
 FT CARBOHYD 375 375  
 FT VARSPPLIC 455 458  
 FT VARSPPLIC 459 521  
 SQ SEQUENCE 521 AA; 57015 MW; 1CB871FAC47D54E CRC64;  
 /FTId=VSP\_002485.  
 /FTId=VSP\_002485.  
 Query Match 7.7%; Score 102; DB 1; Length 521;  
 Best Local Similarity 22.5%; Pred. No. 1.8;  
 Matches 42; Conservative 35; Mismatches 68; Indels 42; Gaps 10;  
 QY 16 LALLPATGQKVKVIGKKGDTVEITCTA-SQKKSIOFHKNNSQIKILNQSGFLTKGDS 74  
 DB 143 ILKPKWITSNNNSNPV-EGDVSILFCDSTYDNDINILYSRN-----GSLFEGD- 191  
 QY 75 KLNDRADSRSLMDQGNFLIKNLKIEDSDRYICEV-----DQKEVQL-LVFG----- 124  
 DB 192 -----RLKISEGNRTLLNLTNRNDTGPVCTRNPSVNRSDPSLNTIYGPDP 242  
 QY 125 LTPNSDTHLLQGSLTLTLTLESPPGSSPSVQC-----RSPRGKNIQGGKTLVSQLELPDS 179  
 DB 243 IISPSDIYIHPGSNLTLSGHA--ASNPPAQYFWLINEXKHASS-----QELFIPNITNNS 296  
 QY 180 GTWTCV 186  
 DB 297 GYTCFV 303  
 RESULT 64  
 VGR2\_RAT STANDARD; PRT; 1343 AA.  
 AC 008775;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular endothelial growth factor receptor 2 precursor (EC 2.7.1.112)  
 DE (VEGFR-2) (Protein-tyrosine kinase receptor flk-1) (Fetal liver kinase  
 DE 1).  
 GN KDR OR FLK1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.



RC TISSUE=Retina;  
 RA Men Y., Edelman J.L., De Vries G.W., Sachs G.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.  
 CC -1- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN  
 CC KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM  
 CC PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF  
 CC VASCULAR PERMEABILITY (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.  
 CC CSD-1/PDGF receptor subfamily.  
 -----  
 CC The SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 CC  
 DR EMBL; U93306; AAB97508.1; -;  
 DR EMBL; U93307; AAB97509.1; -;  
 DR HSSP; P11362; 1FCG.  
 DR InterPro: IPR007110; IG-1-like.  
 DR InterPro: IPR003598; IG-C2.  
 DR InterPro: IPR000719; PrcC\_kinase.  
 DR InterPro: IPR001824; ReceptTyKineIII.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR008266; Tyr\_kinase\_AS.  
 DR Pfam; PF00047; Ig; 6.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 2.  
 DR SMART; SM00408; IGC2; 1.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 5.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 KW Angiogenesis; Transferrase; Tyrosine-protein kinase; Phosphorylation;  
 KW Arg-binding; Receptor; Transmembrane; Signal; Immunoglobulin domain;  
 KW Repeat; Glycoprotein.  
 FT CHAIN 1 19 POTENTIAL.  
 FT 20 1343 VASCULAR ENDOTHELIAL GROWTH FACTOR  
 FT TRANSMEM 760 RECEPTOR 2.  
 FT DOMAIN 782 POTENTIAL.  
 FT DOMAIN 783 1343 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 46 109 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 141 207 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 224 320 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 328 414 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 421 540 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 547 654 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 663 749 IG-LIKE C2-TYPE 7.  
 FT DOMAIN 830 1158 PROTEIN KINASE.  
 FT NP\_BIND 836 844 ATP (BY SIMILARITY).  
 FT BINDING 864 864 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1024 1024 BY SIMILARITY.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 717 717 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD\_RES 1055 1055 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 1343 AA; 150393 MW; AD7E509B62D3FF4 CRC64;  
 Query Match 7.7%; Score 102; DB 1; Length 1343;  
 Best Local Similarity 20.1%; Pred. No. 5.9; Indels 60; Gaps 9;  
 Matches 43; Conservative 37; Mismatches 74;  
 QY 9 HLLVLTALPAATQGNKVKLGKQDVELTCTASQKSIQIFMKNSNQIKLGNQGSF 68  
 DB 542 HVINGPEITVQPAQPTER-----ESMSLLCTADNTEFNLTWYK-----LGSQATS 588  
 QY 69 LTKPSPKLNDRADSRRLM-----DQGNFPIIT--KUKIEDSTIYCEVDQKE- 117  
 DB 589 VHMSES-LTPVCKNLDLWMKLNQTFNSNTDIIIVAFQNASLDQGNVYCSADKKTKK 647  
 QY 118 -----VOLLVGLFRANSDTLLQGSLLTLLESPGSSPSVQGRSPRGRK----- 162  
 DB 648 RHCLVQQLVIERMAPITGNLENQTTI-----GETIEVVC--PTSGNPPLITWFKD 699  
 QY 163 -----IOGKTLVSQLELDQSGTWTG 184  
 DB 700 NETLVDSGLVLDKGNRLTRRVRKEDGGLYTC 733  
 RESULT 65  
 ID SHS1 RAT STANDARD; PRT; 509 AA.  
 AC P97710: 008951; 070426; GQOW15;  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor  
 DE (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-  
 DE regulatory protein alpha-1) (Sitrp-alpha-1) (Brain Ig-1-like molecule  
 DE with tyrosine-based activation motifs) (Blt) (Macrophage fusion  
 DE receptor) (Macrophage membrane protein MFPI50).  
 GN PTPN1 OR SHPS1 OR S1RP OR B1T OR MFR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 42-60; 68-91; 128-137; 150-158;  
 RP 174-189; 192-202; 204-212; 218-237; 259-270; 279-282; 405-415 AND  
 RP 446-453, N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, AND  
 RP INTERACTIONS WITH PTPN6 AND PTPN11.  
 RC TISSUE=Peral fibroblast; PubMed=8943344;  
 RX MEDLINE=97098667; PubMed=8943344;  
 RA Fujioka Y., Matozaki T., Noguchi T., Yamatsu A., Yamao T.,  
 RA Takahashi N., Teuda M., Takada T., Kasuga M.;  
 RA "A novel membrane glycoprotein, SHPS-1, that binds the SH2-domain-  
 RT containing protein tyrosine phosphatase SHP-2 in response to mitogens  
 RT and cell adhesion.";  
 RL Mol. Cell. Biol. 16:6887-6899 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 32-48 AND 446-453, FUNCTION, AND  
 RP PHOSPHORYLATION ON TYROSINE RESIDUES.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=97415431; PubMed=9271230;  
 RA Sano S.-I., Ohnishi H., Omori A., Hasegawa J., Kubota M.;  
 RA "B1T, an immune antigen receptor-like molecule in the brain.";  
 RL FEBS Lett. 411:327-334 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 99-107; 128-149; 197-217; 405-417;  
 RP 419-429; 446-467 AND 496-506, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Rhescher 344; TISSUE=Macrophage;  
 RX MEDLINE=98449911; PubMed=9774638;  
 RA Saginario C., Sterling H., Beckers C., Kobayashi R., Solimena M.,  
 RA Ullu E., Vignery A.;  
 RT "MFR, a putative receptor mediating the fusion of macrophages.";

RA MOJ. Cell Biol. 18:6213-6223(1998).  
 RN [4]  
 RN SEQUENCE OF 1-419 FROM N.A., AND TISSUE SPECIFICITY.  
 RN STRAIN=MAG/RJ1; TISSUE=Alveolar macrophage;  
 RC MEDLINE=98375871; PubMed=97112053;  
 RA Adams S., van der Laan L.J.W., Vernon-Wilson E.,  
 RA Renardel de Lavalette C., Doepp E.A., Dijkstra C.D., Simmons D.L.,  
 RA van den Berg T.K.,  
 RA "Signal-regulatory protein is selectively expressed by myeloid and  
 RT neuronal cells.";  
 RL J. Immunol. 161:1853-1859(1998).  
 RN [5]  
 RN PHOSPHORYLATION IN RESPONSE TO EGF, AND INTERACTION WITH PRPN1.  
 RP MEDLINE=9808865; PubMed=934856;  
 RA Ochi F., Matozaki T., Noguchi T., Fujioke Y., Yamao T., Takeda T.,  
 RA Tenda M., Takeda H., Fukunaga K., Okabayashi Y., Kasuga M.,  
 RT "Epidermal growth factor stimulates the tyrosine phosphorylation of  
 RT SHPS-1 and association of SHPS-1 with SHP-2, a SH2 domain-containing  
 RT protein tyrosine phosphatase.";  
 RL Biochem. Biophys. Res. Commun. 239:483-487(1997).  
 RN [6]  
 RN PHOSPHORYLATION OF TYR-477 AND TYR-501, AND MUTAGENESIS OF TYR-436;  
 RP MEDLINE=99204923; PubMed=9535915;  
 RA Takeda T., Matozaki T., Takeda H., Fukunaga K., Noguchi T.,  
 RA Fujioke Y., Okazaki I., Tenda M., Yamao T., Ochi F., Kasuga M.,  
 RT "Roles of the complex formation of SHPS-1 with SHP-2 in  
 RT insulin-stimulated mitogen-activated protein kinase activation.";  
 RL J. Biol. Chem. 273:9234-9242(1998).  
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts  
 CC as docking protein and induces translocation of PRPN6,  
 CC PRPN1 and other binding partners from the cytosol to the  
 CC plasma membrane. Supports adhesion of cerebellar neurons, neurite  
 CC outgrowth and glial cell attachment. May play a key role in  
 CC intracellular signaling during synaptogenesis and in synaptic  
 CC function. Involved in the negative regulation of receptor tyrosine  
 CC kinase-coupled cellular responses induced by cell adhesion, growth  
 CC factors or insulin. Mediates negative regulation of phagocytosis,  
 CC mast cell activation and dendritic cell activation. CD47 binding  
 CC prevents maturation of immature dendritic cells and inhibits  
 CC cytokine production by mature dendritic cells. May play a role in  
 CC the release of nitric oxide by macrophages. (By similarity).  
 CC -1- SUBUNIT: Binds PRPN1 when tyrosine-phosphorylated, except in  
 CC macrophages, where it primarily binds PRPN6. Binds GNB2 in vitro.  
 CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status  
 CC and forms a stable complex. Binds SCAP1 and/or SCAP2. The  
 CC resulting complex recruits FYN. Binds PRK2 (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain, spleen, lung, liver  
 CC and kidney. Detected at lower levels in heart. Highly expressed in  
 CC alveolar and peritoneal macrophages, and at lower levels in  
 CC dendritic cells.  
 CC -1- PTM: N-glycosylated.  
 CC -1- PTM: Phosphorylated on tyrosine residues in response to insulin,  
 CC cell adhesion or epidermal growth factors. Dephosphorylated by  
 CC PRPN1.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D85183; BAA12734.1; -;  
 CC EMBL: D38468; BAA20368.1; -;  
 CC EMBL: U62328; AAC68478.1; -;  
 CC EMBL: AF055065; AAC18089.1; -;  
 CC HSSP: P01703; 7FAB.  
 CC InterPro: IPR007110; IG-like.

DR	InterPro: IPR003597; IG_C1.	
DR	InterPro: IPR003066; IG_MHC.	
DR	Pfam: Pf00047; Ig; 3.	
DR	SMART: SM00407; IGC1; 2.	
DR	PROSITE: PS50835; IG_LIKE; 3.	
DR	PROSITE: PS00290; IG_MHC; 1.	
KM	Repeat: Signal; Transmembrane; Immunoglobulin domain; SH3-binding;	
KM	Glycoprotein; Phosphorylation.	
FT	SIGNAL	1 31
FT	CHAIN	32 509
FT		
FT	DOMAIN	28 373
FT	TRANSMEM	374 394
FT	DOMAIN	395 509
FT	DOMAIN	32 138
FT	DOMAIN	150 248
FT	DOMAIN	255 349
FT	DISULFID	55 122
FT	DISULFID	172 229
FT	DISULFID	274 332
FT	SITE	436 439
FT	SITE	446 451
FT	SITE	460 463
FT	SITE	477 480
FT	SITE	504 504
FT	MOD_RES	436 436
FT		
FT	MOD_RES	460 460
FT		
FT	MOD_RES	477 477
FT	MOD_RES	501 501
FT	CARBOHYD	54 54
FT	CARBOHYD	93 93
FT	CARBOHYD	169 169
FT	CARBOHYD	181 181
FT	CARBOHYD	205 205
FT	CARBOHYD	209 209
FT	CARBOHYD	242 242
FT	CARBOHYD	246 246
FT	CARBOHYD	271 271
FT	CARBOHYD	293 293
FT	CARBOHYD	312 312
FT	CARBOHYD	320 320
FT	CARBOHYD	345 345
FT	MUTAGEN	436 436
FT		
FT	MUTAGEN	460 460
FT		
FT	MUTAGEN	477 477
FT		
FT	MUTAGEN	501 501
FT		
FT	CONFLICT	8 8
FT	CONFLICT	10 10
FT	CONFLICT	25 25
FT	CONFLICT	58 58
FT	CONFLICT	99 100
FT	CONFLICT	162 162
FT	CONFLICT	189 189
FT	CONFLICT	205 205
FT	CONFLICT	209 209
FT	CONFLICT	405 405
FT	CONFLICT	416 416
FT	CONFLICT	418 421
FT	CONFLICT	450 450

```
FT CONFLICT 499 499 MISSING (IN REF. 3; AA SEQUENCE).
SQ SEQUENCE 509 AA; 55690 MW; 5BE1F0A4DD429F4 CRC64;

Query Match 7.7%; Score 101.5; DB 1; Length 509;
Best Local Similarity 26.0%; Pred. No. 1.9;
Matches 52; Conservative 27; Mismatches 70; Indels 51; Gaps 11;

QY 23 TQGNKVVVGGKKGDVTELTCTASQKRSI-----QFHMKNSNQIKLGNQSGFLTKGSPSKND 78
DB 37 TQADKSVVAAGDSTLNCVSLTPVGPIMFKGSGGNRSFYSFGCHFPRTITVSD 96
QY 79 RADSRRLSDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOLVGLTANSDFHLLQSGS 138
DB 97 -ATKRNMM-----DFEICISNTPEDAGTYIC-VKQKQKIVE-----PDTERKSGCG 141
QY 139 LTLVL---ESP-----PGSSPVPQCR---SPRG--KNIGCGKTLVSQLELQ 177
DB 142 TTLVYLAKPSPEVSGPDSRSGPCQTVNFTCKSYGFSPRNITLKLKDKKEL--SHLE-- 197
QY 178 DSGTWTCTVLONQKVEPKI 197
DB 198 -----TTISSKSNVSYNI 210

RESULT 66
NCA2_MOUSE STANDARD: PRT; 725 AA.
ID_NCA2_MOUSE
AC P13594; Q61950;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM-120)
CN NCAM1 OR NCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=87246524; PubMed=3595563;
RA Barthele D., Santoni M.-J., Wille M., Ruppert C., Calx J.-C.,
RA Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region.";
RL EMO J. 6:907-914 (1987).
RN [2]
RP SEQUENCE OF 20-700 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=89251563; PubMed=2721486;
RA Santoni M.-J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain.";
RL EMO J. 8:385-392 (1989).
RN [3]
RP SEQUENCE OF 642-725 FROM N.A.
RX MEDLINE=88283628; PubMed=3396534;
RA Barthele D., Chaix J.-C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse.";
RL EMO J. 7:625-632 (1988).
RN [4]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401 (1986).
CC -1- FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
```

```
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 120;
CC IsoId=P13594-1; Sequence=displayed;
CC Name=N-CAM 180;
CC IsoId=P13595-1; Sequence=External;
CC Name=N-CAM 140;
CC IsoId=P13595-2; Sequence=External;
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00051; CA68263.1; -
DR EMBL; X15049; CA33148.1; ALT_SEQ.
DR EMBL; X07195; CA30173.1; -.
DR PIR; A29673; IJMSG.
DR PDB; 2NCM; 12-MAR-97.
DR PDB; 3NCM; 23-JUL-99.
DR MGI; MGI:97281; Ncam1.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM0060; FN3; 2.
DR SMART; SM00408; IGC2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
DR Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
DR Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor;
DR 3D-structure.
DR SIGNAL.
DR CHAIN.
DR 1
DR 20
DR 725
DR 111
DR 205
DR 116
DR 302
DR 402
DR 492
DR 519
DR 625
DR 692
DR 152
DR 156
DR 165
DR 161
DR 41
DR 96
DR 139
DR 189
DR 235
DR 288
DR 330
DR 386
DR 427
DR 480
DR 222
DR 316
DR 348
DR 424
DR 450
DR 479
DR 261
DR 268
DR 273
DR 354
DR 355
DR 549
DR 572
DR 575
DR 589
DR 594
DR 602
DR 657
DR 725 AA; 80296 MW; C2AEB8B4461C6B2F CRC64;

Query Match 7.7%; Score 101.5; DB 1; Length 725;
```

Best Local Similarity 19.9%; Pred. No. 3;  
Matches 46; Conservative 42; Mismatches 94; Indels 49; Gaps 10;

QY 3 RG-VPRHLLVLQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHW-KNSNQIK 60  
196 RGEINFDQIVNVPPVQARQSVNATNLGQSVTLVCDADGPFPEPTMSWTKDSEPI 255

QY 61 ILGNQGSFLTKGSPKLNDRAD--SRSLMIOGNFPLIKLKIEDSDTYICEV-----DQ 114  
256 -----NEEDERSRVSQSSSE--VTRRVNDKDEAYVCIENKAGEQ 297

QY 115 KEEVQLLVFG---LTANSDTHLQGGSLTLTLESPGSSPSVQCR-----SPRGNIQ 164  
298 DASHLNVFAKPKITYVENQTMLEEQVTLTCEASDPIPSITWRSTNINISSEQDLD 357

QY 165 GG-----KTLVSQLELDSDGTWC---TVLQNKQKVEFKDIVER 202  
358 GHMVVRSHARVSSLTKSIQYRDAGEVWCTASNTIGDSQSIDLEFGYAPK 408

Db

RESULT 67  
NCAL\_MOUSE STANDARD; PRT; 858 AA.  
ID NCAL\_MOUSE RAT  
AC P13595;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Neural cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)  
DE (NCAM1-140).  
GN NCAM1 OR NCAM.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=8059265; PubMed=3680385;  
RA Small S.O., Shull G.E., Santoni M.-J., Akesson R.;  
RT "Identification of a cDNA clone that contains the complete coding  
sequence for a 140-kD rat NCAM polypeptide.";  
RL J. Cell Biol. 105:2335-2345(1987).  
RN [2]  
RP SEQUENCE OF 355-364 FROM N.A.  
RX MEDLINE=90166485; PubMed=2483093;  
RA Small S.O., Haines S.L., Akesson R.A.;  
RT "Polypeptide variation in an N-CAM extracellular immunoglobulin-like  
fold is developmentally regulated through alternative splicing.";  
RL Neuron 1:1007-1017(1988).  
CC -1- FUNCTION: This protein is a cell adhesion molecule involved in  
neuron-neuron adhesion, neurite fasciculation, outgrowth of  
neurons, etc.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=1;  
Comment=A number of isoforms are produced;  
Name=1;  
Name1: IsoId=P13596-1; Sequence=displayed;  
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
enterprises requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X06564; CAA29809.1; -;  
DR EMBL; M32611; AAA1679.1; -;  
DR PIR; S00846; IJRTNC.  
DR PDB; 1EPF; 27-OCT-00.

DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003598; IG\_C2.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00047; Ig; 5.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00408; IGC2; 5.  
DR PROSITE; PS00835; IG\_LIKE; 5.  
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
KW Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding;  
KW 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 858  
FT DOMAIN 20 721  
FT TRANSMEM 722 739  
FT DOMAIN 740 858  
FT DOMAIN 20 111  
FT DOMAIN 116 205  
FT DOMAIN 212 302  
FT DOMAIN 309 414  
FT DOMAIN 417 502  
FT DOMAIN 514 615  
FT DOMAIN 616 712  
FT DOMAIN 712 712  
FT DOMAIN 152 165  
FT DOMAIN 161 165  
FT DISULFID 41 96  
FT DISULFID 139 189  
FT DISULFID 235 288  
FT DISULFID 330 396  
FT DISULFID 437 490  
FT CARBOHYD 222 222  
FT CARBOHYD 316 316  
FT CARBOHYD 348 348  
FT CARBOHYD 434 434  
FT CARBOHYD 460 460  
FT CARBOHYD 489 489  
SQ SEQUENCE 858 AA; 94658 MW; EALA064AEAO550F6 CRC64;  
Query Match 7.7%; Score 101.5; DB 1; Length 858;  
Best Local Similarity 18.0%; Pred. No. 3.7;  
Matches 43; Conservative 42; Mismatches 99; Indels 55; Gaps 8;

QY 3 RG-VPRHLLVLQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHW-KNSNQIK 60  
196 RGEINFDQIVNVPPVQARQSVNATNLGQSVTLVCDADGPFPEPTMSWTKDSEPI 255

QY 61 ILGNQGSFLTKGSPKLNDRAD--SRSLMIOGNFPLIKLKIEDSDTYICEV-----DQKE 116  
256 -----NEEDDEKHFPSDDSSSLTRVNDKDEAYVCIENKAGEQDA 299

QY 117 EVQLLVFG---LTANSDTHLQGGSLTLTLESPGSSPSVQCR----- 157  
300 SIHLKFAKPKITYVENQTMLEEQVTLTCEASDPIPSITWRSTNINISSEKASWTR 359

QY 158 PRKANIQQG-----KTLVSQLELDSDGTWC---TVLQNKQKVEFKDIVER 202  
360 PEKQETLDGHMVVRSHARVSSLTKSIQYTDAGEVICTASNTIGDSQSVYLEVQYAPK 418

Db

RESULT 68  
NCAL\_MOUSE STANDARD; PRT; 1115 AA.  
ID NCAL\_MOUSE RAT  
AC P13595; Q61949;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Neural cell adhesion molecule 1, 180 kDa isoform precursor (N-CAM 180)  
DE (NCAM1-180).  
GN NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM N-CAM 180).  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=87246524; Pubmed=3595563;  
 RA Barthels D., Santoni M.J., Wille W., Ruppert C., Chaix J.C.,  
 Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;  
 RT "Isolation and nucleotide sequence of mouse NCM cDNA that codes for  
 a Mr 79,000 polypeptide without a membrane-spanning region.";  
 RL EMBO J. 6:907-914(1987).  
 RN [2]  
 RP SEQUENCE OF 529-1115 FROM N.A. (ISOFORM N-CAM 140).  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=88067687; Pubmed=3684567;  
 RA Santoni M.-J., Barthels D., Barbès J.A., Hirsch M.-R., Steinmetz M.,  
 Goridis C., Wille W.;  
 RT "Analysis of cDNA clones that code for the transmembrane forms of the  
 mouse neural cell adhesion molecule (NCAM) and are generated by  
 RT alternative RNA splicing.";  
 RL Nucleic Acids Res. 15:8621-8641(1987).  
 RN [3]  
 RP SEQUENCE OF 642-1115 FROM N.A. (ISOFORM N-CAM 180).  
 RX MEDLINE=88283628; Pubmed=3396534;  
 RA Barbès J.A., Chaix J.C., Steinmetz M., Goridis C.;  
 RT "Differential splicing and alternative polyadenylation generates  
 RT distinct NCAM transcripts and proteins in the mouse.";  
 RL EMBO J. 7:625-632(1988).  
 RN [4]  
 RP SEQUENCE OF 804-1081 FROM N.A. (ISOFORM N-CAM 180).  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=88247737; Pubmed=2454455;  
 RA Barthels D., Vopper G., Wille W.;  
 RT "NCAM-180, the large isoform of the neural cell adhesion molecule of  
 RT the mouse, is encoded by an alternatively spliced transcript.";  
 RL Nucleic Acids Res. 16:4217-4225(1988).  
 RN [5]  
 RP SEQUENCE OF 702-1115 FROM N.A. (ISOFORM N-CAM 140).  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=89251563; Pubmed=2721486;  
 RA Santoni M.J., Barthels D., Vopper G., Boned A., Goridis C., Wille W.;  
 RT "Differential exon usage involving an unusual splicing mechanism  
 RT generates at least eight types of NCAM cDNA in mouse brain.";  
 RL EMBO J. 8:385-392(1989).  
 RN [6]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=86140120; Pubmed=3512556;  
 RA Rougon G., Marshak D.R.;  
 RT "Structural and immunological characterization of the amino-terminal  
 RT domain of mammalian neural cell adhesion molecules.";  
 RL J. Biol. Chem. 261:3396-3401(1986).  
 CC -1- FUNCTION: This protein is a cell adhesion molecule involved in  
 CC neuron-neuron adhesion, neurite fasciculation, outgrowth of  
 CC neurites, etc.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=3;  
 CC Name=N-CAM 180;  
 CC IsoId=P13595-1; Sequence=Displayed;  
 CC Name=N-CAM 140;  
 CC IsoId=P13595-2; Sequence=VSP\_002588;  
 CC Name=N-CAM 120;  
 CC IsoId=P13594-1; Sequence=External;  
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----X07200; CAA30177.1; -;  
 DR EMBL; Y00051; -; NOT ANNOTATED\_CDS.  
 DR EMBL; X06328; CAA29641.1; -;  
 DR EMBL; X07195; CAA30173.1; -;  
 DR EMBL; X07244; CAA30230.1; -;  
 DR EMBL; X15051; CAA33150.1; -;  
 DR EMBL; X15052; CAA33151.1; -;  
 DR PIR; A29673; IJMSNL.  
 DR MGI; 97281; Ncam1.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig\_III.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; fn3; 2.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00060; FN3; 2.  
 DR SMART; SM00408; IGC2; 5.  
 DR PROSITE; PS50835; IG\_LIKE; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
 KW Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1115  
 FT FT  
 FT DOMAIN 20 711 NEURAL CELL ADHESION MOLECULE 1, 180 kDa  
 FT TRANSMEM 712 729 ISOFORM.  
 FT DOMAIN 730 1115 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 20 111 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 116 205 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 212 302 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 309 402 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 407 492 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 519 596 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 625 692 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 152 156 HEPARIN-BINDING (POTENTIAL).  
 FT DOMAIN 161 165 HEPARIN-BINDING (POTENTIAL).  
 FT DISULFID 41 96 PROBABLE.  
 FT DISULFID 139 188 PROBABLE.  
 FT DISULFID 235 289 PROBABLE.  
 FT DISULFID 330 386 PROBABLE.  
 FT CARBOHYD 422 480 PROBABLE.  
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 810 1076 Missing (in isoform N-CAM 140).  
 FT /FTid=VSP\_002588.  
 SQ SEQUENCE 1115 AA; 119351 MW; 2C93DCD474CFBCAF CRC64;  
 Query Match 7.74; Score 101.5; DB 1; Length 1115;  
 Best Local Similarity 19.94; Pred. No. 5.1;  
 Matches 46; Conservative 42; Mismatches 94; Indels 49; Gaps 10;  
 QY 3 RG-VPRPHLLVLTALLPATQGNKVLGKGGDYELTCTAQSQKSIQTHW-KNSQIK 60  
 DB 196 RGEINFDIOVIYVNPPTVQAROSIVNATNLQOSVTLVDAGFPPTMSWTKDGPTE 255  
 QY 61 ILNGQSFLLTKGPKSKNDRAD--SRRLMDQGNPPLIKULKIEDSTYICEV---DQ 114  
 DB 256 -----NEEDERSRSVSUSSE--VTRVNDKNDAEVCIENKAGBQ 297  
 QY 115 KEEOVLVFG---LTRANSOTHLLOQSLTLTLESPGSSPVQCR-----SPRGNIQ 164  
 DB 298 DASHLKVFAPKPKITYVENQTAMELEQVTLTCEASGDPISITWRSTRNISSEQDLD 357  
 QY 165 GG-----KTLVSQLELDQSGTWC---TVLQNKQKVEFKIDIVR 202  
 DB 358 GHWVSRHARVSSLTLSKSIQYRDAGEYMCYASNTIGQDSQSIDLEQYAPK 408  
 RESULT 69

ID	YOR6_ADEG1	STANDARD;	PRT;	135 AA.
AC	P20748:			
DT	01-FEB-1991	(Rel. 17, Created)		
DT	01-FEB-1991	(Rel. 17, Last sequence update)		
DT	01-NOV-1997	(Rel. 35, Last annotation update)		
DE	Hypothetical 15.1 kDa protein (ORF 6) (ORF11).			
OS	Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).			
CC	Viruses; adenoviruses, no RNA stage; Adenoviridae; Aviadenovirus.			
OX	NCBI_TaxID=10553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90251474; PubMed=2160072;			
RA	Akoplian T.A., Kruglyak V.A., Rivkina M.B., Naroditsky B.S.,			
RT	Tikhonenko T.I.;			
RL	"Sequence of an avian adenovirus (CELO) DNA fragment (0-11.2%).";			
RL	Nucleic Acids Res. 18:2825-2825(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96186720; PubMed=8627769;			
RA	Chloocca S., Kurzbauer R., Schaffner G., Baker A., Mautner V.,			
RT	Cotten M.;			
RT	"The complete DNA sequence and genomic organization of the avian			
RT	adenovirus CELO.";			
RL	J. Virol. 70:2939-2949(1996).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on items			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X17217; CAA35086.1; -;			
DR	EMBL; U46933; AAC54932.1; -;			
DR	PIR; S10004; S10004.			
DR	InterPro; IPR007110; Ig-like.			
DR	InterPro; IPR003599; Ig.			
DR	SMART; SM00409; IG; 1.			
KW	Hypothetical protein.			
SC	SEQUENCE 135 AA; 15109 MW; 73871A3D14C56718 CRC64;			
	Query Match	7.7%;	Score 101;	DB 1; Length 135;
	Best Local Similarity	27.5%;	Pred. No. 0.42;	
	Matches 38; Conservative 21; Mismatches 65; Indels 14; Gaps 4;			
Qy	11 LVLVQLALPAT--QGNKVVLGKKGDVVELCTASQKRSIQFHKNSNQIKLGNQSF 68			
Db	1 MLLLVLLVVGTVTLAADHPTLYAPRGSIELGVGAKQKQGVKFEHRFQN-LKIVIAEMSS 59			
Qy	69 LTKGDSKLNDRADRSRLMDQGNFLLIKLKIETSDTYICVEVDQKEEVOLLVGLTAN 128			
Db	60 TNOLEIKRPDNGFQNRSEFPTKHLITTHNASYSDGTSYLHOENDDG-----TEH 110			
Qy	129 SDTH--LLOGOSLTLTLE 144			
Db	111 TDNFKVIVQGSMLTYTLQ 128			
	RESULT 70			
AXOL	AXOL_CHICK	STANDARD;	PRT;	1036 AA.
AC	P28685:			
DT	01-DEC-1992	(Rel. 24, Created)		
DT	01-DEC-1992	(Rel. 24, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Contactin 2 precursor (Axonin-1).			
GN	CNTN2.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archaeosuria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			

RN	NCBI_TaxID=9031;
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Brain;
RX	MEDLINE=92174898; PubMed=1311675;
RA	Zuellig R.A., Rader C., Schroeder A., Kalousek M.B.,
RV	von Bohlen Und Hachsch F., Osterwalder T., Inan G., Stoeckli E.T.,
RT	Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
RT	"The axonally secreted cell adhesion molecule, axonin-1. Primary
RT	structure, immunoglobulin-like and fibronectin-type-III-like domains
RT	and glycocy-1-phosphatidylinositol anchorage.";
RL	Eur. J. Biochem. 204:453-463(1992).
CC	-1- FUNCTION: Axon-associated cell adhesion molecule (AXCAM) which
CC	promotes neurite outgrowth by interaction with the AXCAM L1 (G4)
CC	of neuritic membrane.
CC	-1- SUBCELLULAR LOCATION: Attached to the neuronal membrane by a
CC	GPI-anchor.
CC	-1- PTM: The N-terminus is blocked.
CC	-1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC	-1- SIMILARITY: Contains 4 fibronectin type III domains.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X63101; CAA44815.1; -.
DR	PIR; S22383; S22383.
DR	PDB; 1GS6; 19-MAY-00.
DR	InterPro; IPR008957; FN_III-like.
DR	InterPro; IPR003961; FN_III.
DR	InterPro; IPR007110; IG_Ilike.
DR	InterPro; IPR003598; IG_C2.
DR	Pfam; PF00064; fn3; 3.
DR	SMART; PF00047; Ig; 6.
DR	SMART; SM00060; FN3; 4.
DR	SMART; SM00408; IGC2; 5.
DR	PROSITE; PS50835; IG_LIKE; 6.
KW	Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW	Cell adhesion; Repeat; 3D-structure.
FT	SIGNAL 1 23 OR 25 (POTENTIAL).
FT	CHAIN 24 ? CONTRACTIN 2.
FT	PROPEP ? REMOVED IN MATURE FORM.
FT	DOMAIN 32 123 IG-LIKE C2-TYPE 1.
FT	DOMAIN 128 223 IG-LIKE C2-TYPE 2.
FT	DOMAIN 234 317 IG-LIKE C2-TYPE 3.
FT	DOMAIN 322 406 IG-LIKE C2-TYPE 4.
FT	DOMAIN 412 499 IG-LIKE C2-TYPE 5.
FT	DOMAIN 504 598 IG-LIKE C2-TYPE 6.
FT	DOMAIN 599 608 HINGE (POTENTIAL).
FT	DOMAIN 601 607 GLY/PRO-RICH.
FT	DOMAIN 608 709 FIBRONECTIN TYPE-III 1.
FT	DOMAIN 710 811 FIBRONECTIN TYPE-III 2.
FT	DOMAIN 812 912 FIBRONECTIN TYPE-III 3.
FT	DOMAIN 913 1009 FIBRONECTIN TYPE-III 4.
FT	CARBOHYD 71 71 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 199 199 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 456 456 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 472 472 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 493 493 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 520 520 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 770 770 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 900 900 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 914 914 N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ	SEQUENCE 1036 AA; 113301 MW; 08B80143BE779794 CRC64;
Query Match	7.7%; Score 101; DB 1; Length 1036;
Beet Local Similarity	23.2%; Pred. No. 5.1;
Matches 53; Conservative 30; Mismatches 87; Indels 58; Gaps 9	

```

Oy 34 GDTVELTCTASQKSIQFMWKNNOIKILNGSFLTKGSKLNRADSRSLMDQGNP 93
Db 249 GOMVTLKECFARCONPVQIKMR-----KLDGQ-----TSKMLSSRPL 285
Oy 94 LIIRKLKEDSDTYICEVEDOK---EEVQLIVFG---LRNSPTHLQGSLLTLES 145
Db 286 LHIQNVDEDEBGTVECEAENIKGRDYGRIIHHQPMVLVITDTEADISDLRWSCVA 345
Oy 146 PGSSPSYQ-----CRSPRGKNIQGGKTLVSQLELODSGTWCTYLQNOKV----- 193
Db 346 SGKPPAPARMLRDGQPLASQNRREVSGS-LRFSCLVEDSGMTCQVAENKRGTVYASAE 404
Oy 194 -----EFKIDIVPRASALPAPPTGSALPDPPQTASALPDPPASAL 233
Db 405 LTVQALAPDFRLNPVKR--LIPARNSGKVIIPCQBRAR-----PRATVL 446

RESULT 71
UN52_CABEEL STANDARD: PRT: 3375 AA.
AC 006561: 018261: 018263: 09XTD2: 09XTI5;
DT 01-JUN-1994 (Rel. 29, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane proteoglycan precursor (Perlecan homolog)
DE (Uncoordinated protein 52).
GN UNC-52 OR ZC101.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOPROM A), AND FUNCTION.
RX MEDLINE=9339574; PubMed=893416;
RA RogalSKI T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous
RT to the core protein of the mammalian basement membrane heparan
RT sulfate proteoglycan."
RL Genes Dev. 7:1471-1484(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Percy C.M., Baynes C.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Durbin R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probable role in myofibril assembly and/or attachment
CC of the myofibril lattice to the cell membrane. May be an
CC extracellular anchor for integrin receptors in muscle.
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=;
CC Isoform=Q06561-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=;
CC Isoform=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC Name=;
CC Isoform=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC Note=No experimental confirmation available;
CC Name=;
CC Isoform=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
CC VSP_007196;
CC -1- TISSUE SPECIFICITY: Found in the basement membrane of all
CC contractile tissues. It is concentrated over muscle dense bodies
CC and M-lines which are associated with beta-integrin.
CC -1- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 7 laminin EGF-like domains.

```

```

CC -1- SIMILARITY: Contains 3 laminin G-like domains.
CC -1- SIMILARITY: Contains 2 laminin IV domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL: L13458; AAA28156.1; -
DR EMBL: Z93375; CAB07567.1; -
DR EMBL: Z93395; CAB07567.1; JOINED.
DR EMBL: Z93375; CAB07568.1; -
DR EMBL: Z93395; CAB07568.1; JOINED.
DR EMBL: Z93375; CAB07569.1; -
DR EMBL: Z93395; CAB07569.1; JOINED.
DR EMBL: Z93395; CAB07704.1; -
DR EMBL: Z93395; CAB07706.1; -
DR EMBL: Z93375; CAB07706.1; JOINED.
DR EMBL: Z93395; CAB07707.1; -
DR EMBL: Z93375; CAB07707.1; JOINED.
DR EMBL: Z93375; CAB07708.1; -
DR EMBL: Z93375; CAB07708.1; JOINED.
DR HSSP: P01130; ILDR.
DR WormPep: ZC101.2a; CE15028.
DR WormPep: ZC101.2b; CE15030.
DR WormPep: ZC101.2c; CE15034.
DR WormPep: ZC101.2e; CE18424.
DR GO: GO:0005578; C:extracellular matrix; IEP.
DR GO: GO:0030239; P:myofibril assembly; IEP.
DR InterPro: IPR008985; CoNA_Like_1ec_g1.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF_Like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR007110; IG_Like.
DR InterPro: IPR003599; IG.
DR InterPro: IPR003598; IG_C2.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR002172; LDL_receptor_A.
DR Pfam: PF00047; Ig_16.
DR Pfam: PF00052; Laminin_B_2.
DR Pfam: PF00053; Laminin_EGF_5.
DR Pfam: PF00057; Id1_recept_a; 3.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRODOM: PD003031; Laminin_B_2.
DR SMART: SM00181; EGF_6.
DR SMART: SM00179; EGF_CA_2.
DR SMART: SM00180; EGF_Lam; 6.
DR SMART: SM00409; IG_17.
DR SMART: SM00408; IGC2_17.
DR SMART: SM00281; Lamb; 1.
DR SMART: SM00282; Lamb; 3.
DR SMART: SM00192; LDLa; 3.
DR PROSITE: PS00022; EGF_1; 7.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS00026; EGF_3; 3.
DR PROSITE: PS00835; IG_Like; 17.
DR PROSITE: PS00025; LAM_G_DOMAIN; 3.
DR PROSITE: PS01246; LAMININ_TYPE_EGF; 7.
DR PROSITE: PS01209; LDLRA_2; 3.
DR PROSITE: PS00068; LDLRA_1; 3.
KW Proteoglycan. Glycoprotein. Signal. Immunoglobulin domain. Repeat.
KW Basement membrane. Extracellular matrix. Alternative splicing.
KW Laminin EGF-like domain.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN.
FT DOMAIN 45 130 IG-LIKE C2-TYPE 1.
FT DOMAIN 148 184 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 189 225 LDL-RECEPTOR CLASS A 2.

```

FT	DOMAIN	232	269	LDL-RECEPTOR CLASS A 3.
FT	DOMAIN	271	355	IG-LIKE C2-TYPE 2. (INCOMPLETE).
FT	DOMAIN	384	431	LAMININ EGF-LIKE 1. (INCOMPLETE).
FT	DOMAIN	432	441	LAMININ EGF-LIKE 2 (N-TERMINAL).
FT	DOMAIN	442	633	LAMININ DOMAIN IV 1.
FT	DOMAIN	634	666	LAMININ EGF-LIKE 2 (C-TERMINAL).
FT	DOMAIN	674	720	LAMININ EGF-LIKE 3 (INCOMPLETE).
FT	DOMAIN	721	730	LAMININ EGF-LIKE 4 (N-TERMINAL).
FT	DOMAIN	731	921	LAMININ DOMAIN IV 2.
FT	DOMAIN	922	954	LAMININ EGF-LIKE 4 (C-TERMINAL).
FT	DOMAIN	955	1004	LAMININ EGF-LIKE 5.
FT	DOMAIN	1011	1060	LAMININ EGF-LIKE 6.
FT	DOMAIN	1061	1111	LAMININ EGF-LIKE 7.
FT	DOMAIN	1126	1222	IG-LIKE C2-TYPE 3.
FT	DOMAIN	1226	1311	IG-LIKE C2-TYPE 4.
FT	DOMAIN	1319	1401	IG-LIKE C2-TYPE 5.
FT	DOMAIN	1410	1499	IG-LIKE C2-TYPE 6.
FT	DOMAIN	1503	1585	IG-LIKE C2-TYPE 7.
FT	DOMAIN	1588	1680	IG-LIKE C2-TYPE 8.
FT	DOMAIN	1690	1785	IG-LIKE C2-TYPE 9.
FT	DOMAIN	1793	1878	IG-LIKE C2-TYPE 10.
FT	DOMAIN	1886	1970	IG-LIKE C2-TYPE 11.
FT	DOMAIN	1973	2069	IG-LIKE C2-TYPE 12.
FT	DOMAIN	2073	2163	IG-LIKE C2-TYPE 13.
FT	DOMAIN	2173	2260	IG-LIKE C2-TYPE 14.
FT	DOMAIN	2263	2343	IG-LIKE C2-TYPE 15.
FT	DOMAIN	2349	2435	IG-LIKE C2-TYPE 16.
FT	DOMAIN	2446	2530	IG-LIKE C2-TYPE 17.
FT	DOMAIN	2532	2713	LAMININ G-LIKE 1.
FT	DOMAIN	2793	2960	LAMININ G-LIKE 2.
FT	DOMAIN	2961	3093	GLU-RICH.
FT	DOMAIN	2972	3065	THR-RICH.
FT	DOMAIN	3180	3359	LAMININ G-LIKE 3.
FT	DISULFID	66	114	BY SIMILARITY.
FT	DISULFID	149	161	BY SIMILARITY.
FT	DISULFID	156	174	BY SIMILARITY.
FT	DISULFID	168	183	BY SIMILARITY.
FT	DISULFID	190	202	BY SIMILARITY.
FT	DISULFID	197	215	BY SIMILARITY.
FT	DISULFID	209	224	BY SIMILARITY.
FT	DISULFID	233	246	BY SIMILARITY.
FT	DISULFID	240	259	BY SIMILARITY.
FT	DISULFID	253	268	BY SIMILARITY.
FT	DISULFID	955	964	BY SIMILARITY.
FT	DISULFID	957	971	BY SIMILARITY.
FT	DISULFID	974	983	BY SIMILARITY.
FT	DISULFID	986	1002	BY SIMILARITY.
FT	DISULFID	1011	1021	BY SIMILARITY.
FT	DISULFID	1013	1027	BY SIMILARITY.
FT	DISULFID	1030	1039	BY SIMILARITY.
FT	DISULFID	1042	1058	BY SIMILARITY.
FT	DISULFID	1061	1069	BY SIMILARITY.
FT	DISULFID	1063	1079	BY SIMILARITY.
FT	DISULFID	1082	1091	BY SIMILARITY.
FT	DISULFID	1094	1109	BY SIMILARITY.
FT	DISULFID	1152	1200	BY SIMILARITY.
FT	DISULFID	1338	1384	BY SIMILARITY.
FT	DISULFID	1435	1481	BY SIMILARITY.
FT	DISULFID	1527	1573	BY SIMILARITY.
FT	DISULFID	1618	1663	BY SIMILARITY.
FT	DISULFID	1719	1767	BY SIMILARITY.
FT	DISULFID	1814	1861	BY SIMILARITY.
FT	DISULFID	1907	1954	BY SIMILARITY.
FT	DISULFID	1998	2053	BY SIMILARITY.
FT	DISULFID	2099	2147	BY SIMILARITY.
FT	DISULFID	2195	2242	BY SIMILARITY.
FT	DISULFID	2284	2329	BY SIMILARITY.
FT	DISULFID	2374	2420	BY SIMILARITY.
FT	DISULFID	2467	2514	BY SIMILARITY.
FT	CARBOHYD	1422	1422	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2476	2476	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2950	2950	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3143	3143	N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match      7.7%; Score 101; DB 1; Length 3375;
Best Local Similarity 18.6%; Pred. No. 22;
Matches 50; Conservative 37; Mismatches 106; Indels 76; Gaps 10;

QY      2  NNGVPEFHLLVLTALPAATGKNKVVVGKKGDTVELCTASOKKSIOFWKNSNOIKI 61
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      1581  NRPVSNPARVIVKSPKIRIIPAEQTV--PEGSPKIRKCYVPGHPSVGLTRRRVS----- 1634

QY      62  LKNGQSEFLTKGPKNDRAISRRLMDGNPFLIKNKLEDSPTYICEVD-----Q 114
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      1635  -----GQLNEDADENNGN-----LAVQRAELTDEGDYICTARDPDTCAPID 1675

QY      115  KEEVQLVVGGLVANSDTHTLQSGSLTLTLES---PPGSSPVQCRSP----- 158
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      1676  SPATVHTVNAAPPOVEARPPHVPYITPQTITPEGDPAIQCIVYCGNPSAAQHLSFER 1735

QY      159  -RGKNIQGGKT-----LSVSOLELQDSGTWCTVLQNKKEVEFKIDVPRASALPAPPTG 212
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      1736  VDGKGLPFSSSDRGVLTIPSTQLQDAGEVCLY-----SPENSPPVK 1778

QY      213  SALPDPTQSALPD--PPASALPALAV 239
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      1779  T---NPSTLNVTPEGTPRPVATPPULSV 1804

RESULT 72
KV5Q MOUSE
ID KV5Q MOUSE STANDARD; PRT; 108 AA.
AC P01650;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
OS Ig kappa chain V-V region UPC 61.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE.
RA MEDLINE=79195288; PubMed=109517;
RX Vrana M., Rudikoff S., Potter M.;
RT "The structural basis of a hapten-inhibitable kappa-chain idiotype.";
RL J. Immunol. 123:1905-1910 (1979).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOLETTIES (INULIN).
DR PIR; A92808; KWS61.
DR HSSP; P80362; 1MTL.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SMO0406; IgV_1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 FRAMEWORK-2.
FT DOMAIN 35 49 FRAMEWORK-3.
FT DOMAIN 50 56 FRAMEWORK-4.
FT DOMAIN 57 88 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-4.
FT DOMAIN 98 108 FRAMEWORK-4.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108
SQ SEQUENCE 108 AA; 11809 MW; FAEDADA36076F2AFE CRG64;

Query Match      7.6%; Score 100.5; DB 1; Length 108;
Best Local Similarity 39.0%; Pred. No. 0.34;
Matches 30; Conservative 5; Mismatches 37; Indels 5; Gaps 2;

QY      34  GDTVELCTASOKKSIOFWKNSNOIKIKNQSGFLTKGPKNDRAISRRLMDQG-NF 92
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      16  GDIVTWTQASQGSITNLWFOQK-----GKAPKLIYGASITLEDVPSRFSGRYGTDF 71

```



OY 93 PLINKLIKESDITYC 109  
DB 72 TLTIISLEDEDMATFC 88

RESULT 73  
OPCM\_CHICK STANDARD; PRT; 337 AA.

AC 098892;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Ovid binding protein/cell adhesion molecule homolog precursor  
GN (Neurite inhibitor GP55-A) (OBAM protein gamma isoform).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Lodge A.P., Kim D.-S., Howard M.R., McNamee C.J., Smith N., Moss D.J.;  
RT "Cloning of CEPU-8, a secreted isoform of CEPU-1, and OBAM CDNA from  
RT chick: structural diversity of IGLON family proteins.";  
RL Submitted (Apr-1999) to the EMBL/GenBank/DBD databases.  
RN [2]  
RP SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RA MEDLINE=97157768; PubMed=9004047;  
RA Miledon D.J.A., Kim D.-S., Clarke G.A., Marshall-Clarke S., Moss D.J.;  
RT "A family of glycoproteins (GP55), which inhibit neurite outgrowth,  
RT are members of the Ig superfamily and are related to OBAM,  
RT neurotrophin, LAMP and CEPU-1.";  
RL J. Cell Sci. 109:3129-3138(1996).  
CC -1- FUNCTION: Inhibits neurite outgrowth.  
CC -1- SUBCELLULAR LOCATION: Attached to the nervous system.  
CC -1- TISSUE SPECIFICITY: Restricted to the nervous system.  
CC -1- DEVELOPMENTAL STAGE: Increases during development from very low  
CC levels at embryonic day 10 and is most abundant after hatching.  
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON  
CC family.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Y08170; CAB41420.1; -  
CC InterPro: IPR007110; IG-1like.  
CC InterPro: IPR003598; IG\_c2.  
CC Pfam; PF00047; Ig\_3.  
CC SMART; SM00408; IGC2; 2.  
CC PROSITE; PS50835; IG\_LIKE; 3.  
CC Immunoglobulin domain; Cell adhesion; Glycoprotein; Lipoprotein;  
CC Membrane; GPI-anchor; Repeat; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 314  
FT BY SIMILARITY.  
FT OPIOID BINDING PROTEIN/CELL ADHESION  
FT MOLECULE HOMOLOG.  
FT PROPER 315 337  
FT DOMAIN 32 119  
FT DOMAIN 129 211  
FT DOMAIN 215 302  
FT DISULFID 50 108  
FT DISULFID 150 194  
FT DISULFID 236 288  
FT LIPID 314 314  
FT (Potential).  
FT GPI-anchor amidated asparagine  
FT (Potential).

FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 337 AA; 36887 MW; BAE71755185651E CRC64;

Query Match 7.6%; Score 100.5; DB 1; Length 337;  
Best Local Similarity 21.5%; Pred. No. 1.4;  
Matches 56; Conservative 38; Mismatches 103; Indels 63; Gaps 11;

OY 4 GVPRHLLIVQLLPAATGKNVYLGKGDVETLTCTASQKSIQFHKNSQIKLG 63  
DB 21 GVPRRS-----GDATFPKAMDN--VTVRQESATLCTVDVDRV-RVAMNRSITLYAG 71  
OY 64 NQGSFLTRKPSKNDNRSDRSRLDQGNFPLINLK-----IEDSDYICGE 112  
DB 72 N-----DKNSIDRVVILSNITQVSIKIHNDVDVDEGPTQSVQ 111  
OY 113 ----DQKEVQLVFG----LTNSDTHLLQGSLTLTLESPPSSPSVQCRSPRKN-- 162  
DB 112 TDNHPKTSRVHLIVQVPPQIVNISDLTVNNGSSVTLMCLAFGRPEPTVTRHLSGKGQ 171  
OY 163 -TGGKTLISVQLFLDQSGTWTCTVLQ----NKKVEFKTIVPRASALPAPPGSALP 216  
DB 172 FVSEDEYLETITGIREQSGEYCSAVNDVAVDVRYKVTNVYPYIS--NAKXTGASVG 229  
OY 217 DPOTASALPDPAPASALPAA 236  
DB 230 QKGILQC-----EASAVPVA 244

RESULT 74  
CEK2\_CHICK STANDARD; PRT; 806 AA.

AC P18460;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Tyrosine kinase receptor CEK2 precursor (EC 2.7.1.112).  
GN CEK2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
CC NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90332672; PubMed=2165604;  
RA Pasquale E.B.;  
RT "A distinctive family of embryonic protein-tyrosine kinase  
RT receptors";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:5812-5816(1990).  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Belongs to the fibroblast growth factor receptor  
CC family.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M35195; AAA4664.1; -  
CC PIR; A35963; A35963.  
CC HSSP; P11362; IFGK.  
CC InterPro: IPR007110; IG-1like.  
CC InterPro: IPR003598; IG\_c2.  
CC InterPro: IPR000719; Prot\_kinase.

```

DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR008266; Tyr_kinase_AS.
DR Pfam: PF00047; Ig; 3.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM000408; IGC2; 3.
DR SMART: SM00219; TYKC; 1.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR Receptor: Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KM Transferrase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KM Repeat; Signal.
FT SIGNAL 1 19
FT CHAIN 20 806
FT DOMAIN 20 364
FT TRANSMEM 365 389
FT DOMAIN 390 806
FT DOMAIN 40 124
FT DOMAIN 150 238
FT DOMAIN 247 349
FT DOMAIN 131 141
FT DOMAIN 466 755
FT NP_BIND 472 480
FT BINDING 502 502
FT ACT_SITE 611 611
FT MOD_RES 642 642
FT DISULFID 61 107
FT DISULFID 170 222
FT DISULFID 269 333
FT CARBOHYD 83 83
FT CARBOHYD 96 96
FT CARBOHYD 118 118
FT CARBOHYD 219 219
FT CARBOHYD 256 256
FT CARBOHYD 288 288
FT CARBOHYD 309 309
FT CARBOHYD 322 322
SQ SEQUENCE 806 AA; 89730 MW; B38B3C6D5F2314B6 CRC64;

Query March 7.6%; Score 100.5; DB 1; Length 806;
Best Local Similarity 20.8%; Pred 4.1;
Matches 55; Conservative 32; Mismatches 85; Indels 107; Gaps 12;

QY 10 LLLVQLALPAATQG-----NKVVLGKKGDVVELCTASQKKS1 49
Db 11 LCLAAVAGALPAARRGASRGGAAYLNSETAFLBELVFG--SGDTIELSCN--TQSSSV 68

QY 50 QFHW-----KSNQIKITGNQGSFLTKG 72
Db 69 SVFPMKDGIGIAPSNRTHIGQKLKIINVSYSGLYSCKRHSNE--VIGNFTVRVTD5 126

QY 73 PSKLNDRADRSRLMDQGNPLLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANS DTH 132
Db 127 PSSDDDEDDDES-----EDTGVPFMTRPDKKELLLAV--PAAN----- 164

QY 133 LLAGQSLTLTLESPGSSPSVQC-----RSPRGKNIQGGKTLVSQLET-----QDS 179
Db 165 -----TVRRPCAGNPNPTIYWLKNGKEFGKEHRIQKILHQQMSLVMSVPSDR 217

QY 180 GTWTCYVLQNOQKYE--FKDIVRASALPAPPGSALPDPOT 220
Db 218 GNYTCVVENKXKGNIRHTYQDLVLEERSPHRPIIAG--LPANQT 258

RESULT 75
CEAL_HUMAN STANDARD; PRT; 526 AA.
ID CEAL_HUMAN
AC P13688;
DT 01-JUN-1990 (Rel. 13, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 1 precursor
DE (Biliary glycoprotein 1) (BGP-1) (Antigen CD66) (CD66a antigen).
GN CEACAM1 OR BGP OR BGP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_Taxid:9606;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88320555; PubMed=2457922;
RA Hnoda Y., Neumaier M., Hefta S.A., Drzeniek Z., Wagener C.,
RA Shively L., Hefta L.J.F., Shively J.E., Paxton R.J.;
RT "Molecular cloning of a cDNA coding biliary glycoprotein I: primary
RT structure of a glycoprotein immunologically crossreactive with
RT carcinoembryonic antigen."
RT Proc. Natl. Acad. Sci. U.S.A. 85:6959-6963(1988).
RL [2]
RL REVISION.
RA Hnoda Y., Neumaier M., Hefta S.A., Drzeniek Z., Wagener C.,
RA Shively L., Hefta L.J.F., Shively J.E., Paxton R.J.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:1668-1668(1989).
RL [3]
RL SEQUENCE FROM N.A.
RX MEDLINE=89139550; PubMed=2537311;
RA Barnett T.R., Kreschmer A., Austen D.A., Goebel S.J., Hart J.T.,
RA Biting J.J., Kamatich M.E.;
RT "Carcinoembryonic antigens: alternative splicing accounts for the
RT multiple mRNAs that code for novel members of the carcinoembryonic
RT antigen family."
RL J. Cell Biol. 108:267-276(1989).
RL [4]
RL SEQUENCE FROM N.A.
RP TISSUE=Leukocyte;
RX MEDLINE=91222218; PubMed=2025273;
RA Kuraki M., Arakawa F., Matsuo Y., Oikawa S., Nakazato H.,
RA Matsuo Y.;
RT "Three novel molecular forms of biliary glycoprotein deduced from
RT cDNA clones from a human leukocyte library."
RL Biochem. Biophys. Res. Commun. 176:578-585(1991).
RL [5]
RL SEQUENCE FROM N.A.
RA Lameddin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garne J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Altix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coeffield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoform A).
CC -1- SECRETED (isoforms G, H and I).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=A;
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=P13688-1; Sequence=Displayed;
CC Name=G;
CC IsoId=P13688-2; Sequence=VSP_002482; VSP_002483;
CC Name=H;
CC IsoId=P13688-3; Sequence=VSP_002478; VSP_002479;
CC Name=I;
CC IsoId=P13688-4; Sequence=VSP_002480; VSP_002481;
CC -1- DISEASE: Increased serum levels of BGP-1 are found in individuals
CC suffering from hepatic disorders.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA
CC family.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide C066a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/ncbi/prow/cd/c066a.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```



RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,  
 RA Carriano A.V.;  
 RA Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RX MEDLINE=22398257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Hateshul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Rattellefield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RP [5]  
 RP MEDLINE=98102450; PubMed=9427723;  
 RA Bades-Perner A., Thompson J., van der Putten H., Zimmermann W.;  
 RT "Wice transgenic for the human CGM6 gene express its product, the  
 RT granulocyte marker CD66b, exclusively in granulocytes."  
 RL Blood 91:663-672(1998).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: Expressed in leukocytes of chronic myeloid  
 CC leukemia patients and bone marrow.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.  
 CC -1- DATABASE: NMB=PROV; NOTE=CD guide CD66b entry;  
 CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd66b.htm".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/  
 CC or send an email to license@ebi.ac.uk).  
 CC  
 CC -----  
 CC EMBL: X52378; CAA36604.1; -  
 CC EMBL: M33326; AAA59914.1; -  
 CC EMBL: D90064; BAA14108.1; -  
 CC EMBL: AC004558; AAC13659.1; -  
 CC EMBL: BC026263; AAR26263.1; -  
 CC EMBL: Z95119; CAB08298.1; -  
 CC PIR: S13524; A34815.  
 CC DR Genew: HGNC:1820; CEACAM8.  
 CC MIM: 114890; -  
 CC GO: GO:0005615; C:extracellular space; TAS.  
 CC GO: GO:0005887; C:integral to plasma membrane; TAS.  
 CC GO: GO:0006955; P:immune response; TAS.  
 CC InterPro: IPR007110; IG-1like.  
 CC Pfam: PF00047; IG; 3.  
 CC PROSITE: PS50835; IG LIKE; 2.  
 CC Immunoglobulin domain; Antigen; Signal; Glycoprotein; GPI-anchor;  
 CC Repeat; Polymorphism; Lipoprotein.  
 CC FT SIGNAL 1 34  
 CC CHAIN 35 320 CARCINOEMBRYONIC ANTIGEN-RELATED CELL  
 CC FT PROPEP 321 349 ADHESION MOLECULE 8.  
 CC FT LIPID 320 320 REMOVED IN MATURE FORM.  
 CC FT GPI-anchor amidated aspartate.

FT DOMAIN 35 142 IG-LIKE V-TYPE.  
 FT DOMAIN 145 232 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 237 319 IG-LIKE C2-TYPE 2.  
 FT DISULFID 167 215 PROBABLE.  
 FT DISULFID 259 299 PROBABLE.  
 FT CARBOHYD 104 109 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 114 114 R -> K (1n dbSNP:1041997).  
 FT /FTid=VAR 011721.  
 FT FTid=VAR 011721.  
 FT CONFLICT 322 332 L -> V (IN REF. 2).  
 FT SEQ 349 AA; 38154 MM; AACF74DA1AC839D8 CRC64;  
 SQ  
 Query Match 7.5%; Score 98.5; DB 1; Length 349;  
 Best Local Similarity 20.2%; Pred. No. 2.1;  
 Matches 60; Conservative 51; Mismatches 87; Indels 99; Gaps 15;  
 QY 5 VPPRHLLVQL-----ALPAAIGKNTVL-----GKKDIT 36  
 DB 13 IPQGLLTASLFTFWNPPTTAQLTEAVPSNAEKEVLLVHNLPPDPRGYMYKGET 72  
 QY 37 VELTASQKSGIOWFKNSNOIKLGNOSFLLTKGPSKLNRAADRSRLMDGNPPLIT 96  
 DB 73 VD-----ANRRITIG-----VYSNOQ--ITPSPANS-----RETIYP--NLSLM 109  
 QY 97 KKLKIEDSDTY---ICEVEDQKEV---QLLVF-----GLTANSDFHLLQGSLLTLES 145  
 DB 110 RVNTRDTSYTLQVILKMLMSEEVYQGPSVHPETPKPSISSNNSNPVEDKDAVFTCEP 169  
 QY 146 PP-----GSSPVQCRSPRGKNIQGGKTLVSQLELDSGTWTCTVLOKQKVER 195  
 DB 170 ETQNTTYLWVNGQSLPV--SPRLQLSNGNRTLLSTVRDVGVECEI-ON----- 219  
 QY 196 KIDIVPRASALPAPPTGSALPPQRTASALPDPAPASALPALVALVISFLGLGIVAC 252  
 DB 220 -----PASANFSDPVTNLVXGPDAPITSD--TYHAGVNLNLSC 259  
 RESULT 77  
 SLIB\_HUMAN STANDARD; PRT; 686 AA.  
 AC Q96RT6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Static acid binding Ig-like lectin 11 precursor (Siglec-11) (Static  
 DE acid-binding lectin 11) (UNQ9222/PRO28718).  
 GN SIGLEC11.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH PTPN6 AND PTPN11.  
 RX MEDLINE=22086217; PubMed=11986377;  
 RA Angata T., Kerr S.C., Greaves D.R., Varki N.M., Crocker P.R.,  
 RA Varki A.;  
 RT "Cloning and characterization of human Siglec-11. A recently evolved  
 RT signaling that can interact with SHP-1 and SHP-2 and is expressed by  
 RT tissue macrophages, including brain microglia."  
 RT J. Biol. Chem. 277:24466-24474(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=22887296; PubMed=12975309;  
 RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,

RA	Chen J., Chow B., Chui C., Crowley C., Currell B., Devel B., Dowd P.,
RA	Hacon D., Foster J., Grimaldi C., Gu Q., Haas P.E., Haldens S.,
RA	Huang A., Kim H.S., Klimowaki L., Jin Y., Johnson S., Lee J.,
RA	Lewis L., Liao D., Mark M., Robble E., Sanchez C., Schoenfeld J.,
RA	Seeshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA	Vandlen R., Weanabe C., Wleand D., Woods K., Xie M.-H., Yanura D.,
RA	Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA	Goddard P.,
RT	"The secreted protein discovery initiative (SPDI), a large-scale
RT	effort to identify novel human secreted and transmembrane proteins: a
RT	bioinformatics assessment."
RL	Genome Res. 13:2265-2270 (2003).
CC	-1- FUNCTION: Putative adhesion molecule that mediates sialic-acid
CC	dependent binding to cells. Preferentially binds to alpha2,8-
CC	linked sialic acid. The sialic acid recognition site may be masked
CC	by cis interactions with sialic acids on the same cell surface. In
CC	the immune response, may act as an inhibitory receptor upon ligand
CC	induced tyrosine phosphorylation by recruiting cytoplasmic
CC	phosphatase(s) via their SH2 domain(s) that block signal
CC	transduction through dephosphorylation of signaling molecules.
CC	-1- SUBUNIT: Interacts with PTPN6/SHP-1 and PTPN11/SHP-2 upon
CC	phosphorylation.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event-Alternative splicing; Named isoforms=2;
CC	Name=1;
CC	IsoId=Q96RL6-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId=Q96RL6-2; Sequence=VSP_008764;
CC	-1- TISSUE SPECIFICITY: Expressed by macrophages in various tissues
CC	including Kupffer cells. Also found in brain microglia.
CC	-1- DOMAIN: Contains 1 copy of a cytoplasmic motif that is referred
CC	to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
CC	This motif is involved in downmodulation of cellular responses.
CC	The phosphorylated ITIM motif binds to the SH2 domain of
CC	PTPN6/SHP-1.
CC	-1- PTM: Phosphorylated on tyrosine residues.
CC	-1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC	(sialic acid binding Ig-like lectin) family.
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC	-1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> )
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF337818; AAK72907.1; -
DR	EMBL; AY358135; AAO88502.1; -
DR	Genew; HGNC:15622; SIGLEC11.
DR	MIM; 607157;
DR	InterPro; IPR007110; IG-1-like.
DR	InterPro; IPR003598; IG_C2.
DR	InterPro; IPR003006; IG_MHC.
DR	Pfam; PF00047; Ig_3.
DR	SMART; SMO0408; IGC2; 2.
DR	PROSITE; PSS0835; IG_LIKE; 3.
DR	PROSITE; PSS00290; IG_MHC; 1.
KW	Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;
KW	Lectin; Glycoprotein; Phosphorylation; Alternative splicing.
FT	CHAIN
FT	SIGNAL
FT	1
FT	15
FT	STATIC ACID BINDING IG-LIKE LECTIN 11.
FT	DOMAIN
FT	16
FT	549
FT	TRANSMEM
FT	550
FT	572
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).
FT	IG-LIKE V-TYPE.
FT	DOMAIN
FT	19
FT	122
FT	IG-LIKE C2-TYPE 1.
FT	DOMAIN
FT	147
FT	232
FT	IG-LIKE C2-TYPE 2.
FT	DOMAIN
FT	239
FT	338
FT	IG-LIKE C2-TYPE 3.
FT	DOMAIN
FT	343
FT	440
FT	ITIM MOTIF.
FT	SITE
FT	630
FT	635

FT	DISULFID	37	174	BY SIMILARITY.
FT	DISULFID	42	102	BY SIMILARITY.
FT	DISULFID	165	216	BY SIMILARITY.
FT	DISULFID	275	322	BY SIMILARITY.
FT	DISULFID	379	424	BY SIMILARITY.
FT	CARBOHYD	43	43	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	354	354	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	363	363	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	503	503	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	VANSPPLIC	443	539	YPOLGSGFEVTPSSAGCPMANSLSLHGSLGSLRCRCAMN
FT				VHGAQSGSVFQLPG -> W (in isoform 2).
FT				/Fttd=VSP 008764.
FT	CONFLICT	84	84	E -> A (IN REF. 2).
FT	CONFLICT	353	353	A -> G (IN REF. 2).
SEQ	SEQUENCE	686 AA;	74544 MW;	FDDEA1936156558 CRC64;
Query Match				
Best Local Similarity 7.5%; Score 98.5; DB 1; Length 686;				
Matches 61; Conservative 38; Mismatches 99; Indels 97; Gaps 12				
QY	24	QGNKVVLG-KKGDVVELCTHSQKKSIOFHKMNSNQIKLNGQSFITKSGSKLNDKRD	82	
DB	257	QGNVYILEVQKGFRLICADISOPATLSW-----VLQDRVL	295	
QY	83	RRLMDQGNFPLIKLKIEDSDPYICEVQO-----KEEYQLVYF	123	
DB	296	SSHFWGPTTLGLRLGVAGSGRYTCAENRLSGQOALDSVYPPENLRVWSQANR	355	
QY	124	-----GLTANSDTHLLOGSLTL-----TLESPPGSSPSVQCRSPRKNIOGKTLVSQ	173	
DB	356	TVLENLNGGTSFLPYLEGGSLKLVCTHSSPPA-----RLSWTMGQTVGPSQSD	405	
QY	174	-----LEIQDSGTWTCTV-----LQONK-KVEFIDIVPR-----A	203	
DB	406	PGVLELPEIQMEHSEFPTCHQHPRLGSGHVSLSLVHPPLDLPGPSCEWAEGLHCSS	465	
QY	204	SALPAPP-----TGSALPDPTQASALPDPPASALPAPALAVISFLGLT--GLGVAC	252	
DB	466	QASPPSLRWLGEELLGNSSQGSFEVTPSSAGPMANSLSLHGSLGSLRLRC	520	
RESULT 78				
CD8B_SAI	CD8B_SAI	STANDARD;	PRT;	209 AA.
AC	Q9XSW7			
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	T-cell surface glycoprotein CD8 beta chain precursor (Antigen CD8B).			
GN	CD8B.			
OS	Saimiri sciureus (Common squirrel monkey).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.			
OX	NCBI_TaxID=9521;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=92039;			
RX	MEDLINE=99299362; PubMed=10369934;			
RA	Ureca-Vidal A., Garcia Z., Lemonnier F.A., Kazanji M.;			
RT	"Molecular characterization of cDNAs encoding squirrel monkey (Saimiri			
RL	sciureus) CD8 alpha and beta chains.";			
CC	Immunogenetics 49:718-721(1999).			
CC	-1- FUNCTION: Identifies cytotoxic/suppressor T-cells that interact			
CC	with MHC class I bearing targets. CD8 is thought to play a role in			
CC	the process of T-cell mediated killing.			
CC	-1- SUBUNIT: In general heterodimer of an alpha and a beta chain			
CC	linked by two disulfide bonds.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.			

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ130819; CAB41463.1; -
DR GO; GO:0042101; C-T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042288; F:MHC class I protein binding; ISS.
DR GO; GO:0005515; F:protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0042110; P:T-cell activation; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003599; Ig.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00409; Ig; 1.
DR POSITE; PS50835; IG LIKE; 1.
KM Immunoglobulin domain; Transmembrane; T-cell; Antigen; Glycoprotein;
KW Immune response; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 209 T-CELL SURFACE GLYCOPROTEIN CD8 BETA
FT DOMAIN 22 169 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 170 190 POTENTIAL.
FT DOMAIN 191 209 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 122 IG-LIKE V-TYPE.
FT DISULFID 41 115 POTENTIAL.
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 209 AA; 23481 MW; E5958E39DF629C5 CRC64;

Query March 7.4%; Score 98; DB 1; Length 209;
Best Local Similarity 26.1%; Pred. No. 1.2;
Matches 46; Conservative 27; Mismatches 73; Indels 30; Gaps 8;

QY 6 PFRHLVLQALPAATQGNKVV-----VLKKGDTVELTCTASQKSIQFHM----- 53
DB 3 PRMWLLSAQL-----ALHNGSVLQTPAYIMVQTNQMWLSCAIISSSTRIYWLQLH 58
QY 54 --KNSNOIKITGNOSFLTKSPKLNDRADSRRLMOCN-FPIITKLKIEDSDTYICE 110
DB 59 APSSNSHHEILAFWDS--SKGTIHSEGEQKRTVFRDGSILFFLNTLRVXLEDSGYFCM 116
QY 111 VEDQKEVQLLVFGLTANSPTHLIQGSLTLTLESPGSSP-SVQCRSPGKNIQOG 165
DB 117 VIGS-----PTLIFG-----TGTQLSVVDILPTTAQTCKSPKTVCKLPAPETRK 164

RESULT 79
PIGR HUMAN STANDARD; PRT; 764 AA.
AC P01833;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polymetric-immunoglobulin receptor precursor (poly-Ig receptor) (PIGR)
DE [Contains: Secretory component].
GN PIGR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=92039621; PubMed=1682231;
RA Krawiec P., Grzeschik K.H., Geurts van Kessel A.H., Olausen B.,
RA Brandtzaeg P.;
RT "The human transmembrane secretory component (poly-Ig receptor):
RT molecular cloning, restriction fragment length polymorphism and

```

```

RT Chromosomal sublocalization.";
RL Hum. Genet. 87:642-648 (1991).
RN [2]
RX MEDLINE=92387236; PubMed=1355431;
RA Krawiec P., Kvale D., Taeken K., Brandtzaeg P.;
RT "Molecular cloning and exon-intron mapping of the gene encoding human
RT transmembrane secretory component (the poly-Ig receptor).";
RL Eur. J. Immunol. 22:2309-2315 (1992).
RN [3]
RX MEDLINE=89149795; PubMed=2920039;
RA Krawiec P., Solberg R., Sandberg M., Oyen O., Janssen T.,
RA Brandtzaeg P.;
RT "Molecular cloning of the human transmembrane secretory component
RT (poly-Ig receptor) and its mRNA expression in human tissues.";
RL Biochem. Biophys. Res. Commun. 158:783-789 (1989).
RN [4]
RX MEDLINE=85128981; PubMed=6526384;
RA Biffert H., Quentin E., Decker J., Hillemeir S., Hufschmidt M.,
RA Klingmüller D., Weber M.H., Hilschmann N.;
RT "The primary structure of human free secretory component and the
RT arrangement of disulfide bonds.";
RL Hoppe-Seyler's Z. Physiol. Chem. 365:1489-1495 (1984).
RN [5]
RX MEDLINE=91315750; PubMed=1859628;
RA Biffert H., Hilschmann N.;
RT "Determination of the molecular structure of the human free secretory
RT component.";
RL Biol. Chem. Hoppe-Seyler 372:119-128 (1991).
RN [6]
RX MEDLINE=97379357; PubMed=9237679;
RA Hughes G.J., Frutiger S., Savoy L.-A., Reason A.J., Morris H.R.,
RA Jaton J.-C.;
RT "Human free secretory component is composed of the first 585 amino
RT acid residues of the polymetric immunoglobulin receptor.";
RL FEBS Lett. 410:443-446 (1997).
CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC TRANSMEMBRANE SEGMENT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Also secreted.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like V-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S62403; AAB20203.1; -
DR EMBL; S43449; AAB23176.1; JOINED.
DR EMBL; S43437; AAB23176.1; JOINED.
DR EMBL; S43441; AAB23176.1; JOINED.
DR EMBL; S43442; AAB23176.1; JOINED.
DR EMBL; S43443; AAB23176.1; JOINED.
DR EMBL; S43444; AAB23176.1; JOINED.
DR EMBL; S43445; AAB23176.1; JOINED.
DR EMBL; S43446; AAB23176.1; JOINED.
DR EMBL; S43447; AAB23176.1; JOINED.
DR EMBL; S43448; AAB23176.1; JOINED.
DR EMBL; M24559; AAA36102.1; -
DR EMBL; A52091; CAA03384.1; -
DR PIR; A46537; ORHUGS.
DR Glycositedb; P01833; -

```

	DR	MIM; 173880;	-	Genew; HGNC:8968; PIGR.
	DR	GO; GO:0005887;	C:integral to plasma membrane; TAS.	
	DR	InterPro; IPR007110;	ig-like.	
	DR	InterPro; IPR003599;	Ig.	
	DR	Pfam; PF00047;	Ig; 5.	
	DR	SMART; SMO0409;	Ig; 5.	
	DR	PROSITE; PS50835;	IG-LIKE; 2.	
	KW	Immunoglobulin domain;	Repeat; Transmembrane; Glycoprotein; Signal;	
	KM	Polymorphism.		
	FT SIGNAL	.1	18	POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
	FT CHAIN	19	764	SECRETORY COMPONENT.
	FT CHAIN	19	603	EXTRACELLULAR (POTENTIAL).
	FT DOMAIN	19	638	POTENTIAL.
	FT TRANSMEM	639	661	CYTOSOLASMIC (POTENTIAL).
	FT DOMAIN	662	764	IG-LIKE V-TYPE 1.
	FT FT	19	120	IG-LIKE V-TYPE 2.
	FT DOMAIN	145	237	IG-LIKE V-TYPE 3.
	FT FT	250	352	IG-LIKE V-TYPE 4.
	FT DOMAIN	364	458	IG-LIKE V-TYPE 5.
	FT DISULFID	462	561	
	FT DISULFID	40	110	
	FT DISULFID	56	64	
	FT DISULFID	152	220	
	FT DISULFID	257	325	
	FT DISULFID	271	279	
	FT DISULFID	371	441	
	FT DISULFID	385	395	
	FT DISULFID	482	544	
	FT DISULFID	486	520	
	FT DISULFID	496	503	
	FT CARBOHYD	83	83	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	90	90	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	135	135	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	186	186	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	421	421	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	469	469	N-LINKED (GLCNAC . . ) .
	FT CARBOHYD	499	499	N-LINKED (GLCNAC . . ) .
	FT VARIANT	580	580	A -> V.
	FT CONFLICT	136	136	/FTID=VAR_003920.
	FT CONFLICT	158	158	D -> Q (IN REF. 4 AND 5) .
	FT CONFLICT	208	209	N -> D (IN REF. 4 AND 5) .
	FT CONFLICT	229	229	NO -> DE (IN REF. 4 AND 5) .
	FT CONFLICT	229	229	MISSING (IN REF. 4 AND 5) .
	FT CONFLICT	234	234	E -> N (IN REF. 4 AND 5) .
	FT CONFLICT	241	241	H -> O (IN REF. 4 AND 5) .
	FT CONFLICT	262	262	B -> Q (IN REF. 4 AND 5) .
	FT CONFLICT	280	280	D -> N (IN REF. 4 AND 5) .
	FT CONFLICT	392	392	N -> D (IN REF. 4 AND 5) .
	FT CONFLICT	500	500	N -> D (IN REF. 4 AND 5) .
	SQ SEQUENCE	764 AA;	83313 MW;	916B3B662C319950 CRC64;
	Query Match	7.4%;	Score 98;	DB 1; Length 764;
	Best Local Similarity	20.2%;	Fold No. 5.9;	
	Matches	53;	Conservative	36; Mismatches 101; Indels 72; Gaps 10.
Oy	2	NRGVPFRHLILVLTALPAATQGKRVVLGGKDITVELTLC-----TASOKSISGFHWXNS	56	
Dd	116	SRGISFD---VSLEWSGSRGLLNDRTKYTVLDGRVTTCNCPKTMAQRKSLSL-----	165	
Oy	57	NGIKTLGNQSFLTKSPSKLANDRSRLSDOQN----PLLIRKLKIESTDYICEVE	112	
Dd	166	--YKOIGLYPALVIDSSGVVNPNYTGRIRLDIQTGQLLFSSFVINQLRHSLDGAYLCOAG	223	
Oy	113	DKKEVOLLVLGULTNSPDTHLLQQSGSLTTLESPPGS-----SPSYQ-----CRSPR	159	
Dd	224	DDSNRNK-----KNADLOVKLRP--ELUYEDLGSSTFHCAIAPEVANVAKKFCROSS	275	
Oy	160	GKNI-----OGAKTSIVSQLETDLDSGTWTCTV-----ION	189	
Dd	276	GENDCVVVNTLGKRAAFEGRTLNPQRDXGFSFYVIITGKRLEDAGRILCGAHSDGQLDE	335	
Oy	190	QKVVEFKIDIVRASALSAPPPT	211	

```

Db      336 GSP1QAWQLFVNEEST1PRSP 357

RESULT 80
TYO3_MOUSE
ID      TYO3_MOUSE      STANDARD,      PRT,      680 AA.
AC      P55144,
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Tyrosine-protein kinase receptor TYRO3 precursor (EC 2.7.1.112)
DE      (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase DTK) (TK19-2).
GN      TYRO3 OR DTK OR RSE.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10990;
[1]
SEQUENCE FROM N.A.
RP      TISSUE=Brain;
RC      MEDLINE=94193774; PubMed=7511603;
RX      Mark M.R., Seaden D.T., Wang Z., Gu Q., Goddard A., Godowski P.J.;
RT      "RE, a novel receptor-type tyrosine kinase with homology to Axl/Uro,
RT      is expressed at high levels in the brain.";
RL      J. Biol. Chem. 269:10720-10728(1994).
[2]
SEQUENCE FROM N.A.
RP      TISSUE=Brain;
RC      MEDLINE=95161079; PubMed=7857657;
RX      Crosier P.S., Lewis P.M., Hall L.R., Vitas M.R., Morris C.M.,
RA      Beier D.R., Wood C.R., Crosier K.E.;
RT      "Isolation of a receptor tyrosine kinase (DTK) from embryonic stem
RT      cells: structure, genetic mapping and analysis of expression.";
RL      Growth Factors 11:125-136(1994).
[3]
SEQUENCE FROM N.A.
RP      STRAIN=C57BL/6;
RC      MEDLINE=94336210; PubMed=8058320;
RX      Lai C., Gore M., Lemke G.;
RT      "Structure, expression, and activity of Tyro 3, a neural adhesion-
RT      related receptor tyrosine kinase.";
RL      Oncogene 9:2567-2578(1994).
[4]
SEQUENCE FROM N.A.
RP      STRAIN=BALB/c; TISSUE=Brain;
RC      MEDLINE=95240399; PubMed=7723626;
RX      Schulz N., Paulhac C., Lee U., Zhou R.;
RT      "Isolation and expression analysis of tyro3, a murine growth factor
RT      receptor tyrosine kinase preferentially expressed in adult brain.";
RL      Brain Res. Mol. Brain Res. 28:273-280(1995).
[5]
SEQUENCE FROM N.A.
RP      STRAIN=BALB/c; TISSUE=Brain;
RC      Sasaki M.;
RL      Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: May be involved in cell adhesion processes, particularly
CC      in the central nervous system.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: Abundant in the brain and lower levels in
CC      other tissues.
CC      -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC      AXL/URO SUBFAMILY.
CC      -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC      -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; U05683; AAA19237.1; -

DR EMBL; U18933; AAC52148.1; -

DR EMBL; X78103; CAAS4995.1; ALT INIT.

DR EMBL; U18342; AAB26942.1; ALT\_INIT.

DR EMBL; AB000828; BAA19193.1; -

DR PIR; B53743; B53743.

DR HSP; P11362; IFGK.

DR MGJ; MG1.104294; TYRO3.

DR InterPro; IPR008957; FN\_III-like.

DR InterPro; IPR003961; FN\_III.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003598; Ig\_c2.

DR InterPro; IPR007119; Prot\_kinase.

DR InterPro; IPR001245; Tyr\_kinase.

DR InterPro; IPR008266; Tyr\_kinase\_AS.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF00047; Ig; 2.

DR Pfam; PF00069; kinase; 1.

DR PRINTS; PR00109; TYRKINASE.

DR PRODOM; PD000001; Prot\_kinase; 1.

DR SMART; SM00408; IGC2; 1.

DR SMART; SM00219; Tyrc; 1.

DR PROSITE; PS50835; IG-LIKE; 2.

DR PROSITE; PS50107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.

DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.

KW Cell adhesion; Transferrase; Tyrosine-protein kinase; Receptor;

KW ATP-binding; Repeat; Signal; Transmembrane; Immunoglobulin domain;

KW Glycoprotein; Phosphorylation.

FT SIGNAL 1 30

FT CHAIN 1 880

FT DOMAIN 31 419

FT TRANSMEM 420 440

FT DOMAIN 441 880

FT DOMAIN 31 118

FT DOMAIN 129 209

FT DOMAIN 214 301

FT DOMAIN 305 401

FT DOMAIN 508 785

FT NP BIND 514 522

FT BINDING 540 540

FT ACT\_SITE 645 645

FT DISULFID 54 107

FT CARBOHYD 53 193

FT CARBOHYD 75 75

FT CARBOHYD 181 181

FT CARBOHYD 220 220

FT CARBOHYD 230 230

FT CARBOHYD 283 283

FT CARBOHYD 356 356

FT CARBOHYD 370 370

FT MOD\_RES 676 676

FT CONFLICT 630 630

FT CONFLICT 811 811

SEQ SEQUENCE 880 AA; 96222 MW; 387AC36CB016B3F7 CRC64;

Query Match 7.4%; Score 98; DB 1; Length 880;

Best Local Similarity 20.1%; Pred. No. 7;

Matches 59; Conservative 44; Mismatches 97; Indels 94; Gaps 11;

QY 4 GVPRHLLVQLA--LLPATQGNKVLG-----KSGDVELTCTASQKSIQPHW- 53

DB 8 GWPLRLPILLAGLSLLPGSAAGLKLMAFPVMTYSGQGPVTLNCSVEGMEPDTHM 67

QY 54 -----KNSNQIKILGNQGSFLTYGPKSLNDRADSRSLMDQGNFLIKNLKIEDSDTY 107

DB 68 KDTGVVONASQVSISSIEHSGWIGL-----LSLKSVERSDAGLY 105

QY 108 ICEVEDQKEEYOLVGLTANSQDTHLQGSLLTLES-----PGSSPSVQ 154

DB 106 WCVVND-GEETKI-----SQSWLVTEGVFPFVEPKDLAVPNAPFQIS 149

QY 155 CRSP-----RGKNIQGGKTLVSQLE---LQDSGTWCTVLAQNKVEFKIDIV 200

DB 150 CEAVGPPEPVITVMMGLTKVGGPAPSPVLVNTGVTGRTBSPSCA-RNKKGLATSRPAI 208

QY 201 PRASALPAPPPTGALPDPTAGALPDPPASALPALAVISFLIGLGVACVL 254

DB 209 VRLQAPPAPAFNTV-----TTISSYNASVAVWPGADGLALHSTCV 250

RESULT 81

KVST\_MOUSE

ID KVST\_MOUSE STANDARD; PRT; 108 AA.

AC P01653;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Ig kappa chain V-V region W3082.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;

RN [1]

RP SEQUENCE.

RX MEDLINE=82099361; PubMed=6798111;

RA Johnson N., Slinkard J., Paul L., Hood L.;

RT "The complete V domain amino acid sequences of two myeloma inulin-binding proteins."

RT J. Immunol. 128:302-307(1982).

CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).

DR PIR; B92811; KYMS82.

DR HSP; P80362; IWTL.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003596; Ig\_v.

DR Pfam; PF00047; Ig; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG-LIKE; 1.

KW Immunoglobulin V region.

FT DOMAIN 1 23

FT DOMAIN 24 34

FT DOMAIN 35 49

FT DOMAIN 50 56

FT DOMAIN 57 88

FT DOMAIN 89 97

FT DOMAIN 98 108

FT DISULFID 23 88

FT NON\_TER 108 108

SEQ SEQUENCE 108 AA; 11850 MW; C5C145DC376F30CD CRC64;

Query Match 7.4%; Score 97.5; DB 1; Length 108;

Best Local Similarity 37.7%; Pred. No. 0.58;

Matches 29; Conservative 6; Mismatches 37; Indels 5; Gaps 2;

QY 34 GDTVELTCTASQKSIQFHKNSNQIKILGNQGSFLTYGPKSLNDRADSRSLMDQGNF 92

DB 16 GDIVTWTQASQGTININLMWFQKRP---GKAPKLIVGASILEDGVSPRFSGSRYGTD 71

DB 72 TLTISSLEDDMATYFC 88

RESULT 82

ICOL\_MOUSE

ID ICOL\_MOUSE STANDARD; PRT; 322 AA.

AC Q9JH8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)



DE ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)  
DE (B7-related protein-1) (B7RP-1) (LICOS).  
GN ICOSL OR B7H2 OR B7RP1.  
OS Eue musculus (Mouse).  
OC Murayovota; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN  
RN SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
RC TISSUE=Lymphocytes;  
RX Yoshinaga S.K., Whoriskey J.S., Khare S.D., Sarmiento U., Guo J.,  
RA Horan T., Shih G., Zhang M., Cocci M.A., Kohno T., Tafuri-Bladt A.,  
RA Brenan D., Campbell P., Chang D., Chiu L., Dai T., Duncan G.,  
RA Elliott G.S., Hui A., McCabe S.M., Scully S., Shahinian A.,  
RA Shackle C.L., Van G., Mak T.W., Senaldi G.;  
RT "T-cell co-stimulation through B7RP-1 and ICOS".  
RL Nature 402:827-832(1999).  
RN  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Thymus;  
RX MEDLINE=20015817; PubMed=10549624;  
RA Swallow M.M., Wallin J.J., Sha W.C.;  
RT "B7h, a novel costimulatory homolog of B7.1 and B7.2, is induced by  
RT TNFalpha.";  
RL Immunity 11:423-432(1999).  
RN  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=C3H/HeJ; TISSUE=Fetal thymus;  
RX MEDLINE=20126021; PubMed=10567666;  
RA Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,  
RA Leonard J.P., Hunter S.E., Zoller R., Thomas J.L., Miyashiro J.S.,  
RA Jacobs K.A., Collins M.;  
RT "Identification of GL50, a novel B7-like protein that functionally  
RT binds to ICOS receptor.";  
RL J. Immunol. 164:1653-1657(2000).  
RN  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Periphereal blood lymphocytes;  
RX MEDLINE=21286479; PubMed=11390400;  
RA Ling V., Wu P.W., Miyashiro J.S., Marusic S., Finnerty H.F.,  
RA Collins M.;  
RT "Differential expression of inducible costimulator-ligand splice  
RT variants: lymphoid regulation of mouse g150-b and human g150  
RT molecules.";  
RL J. Immunol. 166:7300-7308(2001).  
RN  
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RA Ling V., Dunnwald-Joannopoulos K.;  
RT "G150 molecules and uses thereof.";  
RL Patent number WO0121796, 29-MAR-2001.  
CC  
CC -1- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR  
CC ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND  
CC CYTOKINE SECRETION, INDUCES ALSO B-CELL DIFFERENTIATION AND  
CC DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN  
CC MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS  
CC WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-  
CC STIMULATING MEMORY T CELL FUNCTION. DURING PREGNANCY, MAY FUNCTION  
CC TO SKEW THE CYTOKINE OF MATERNAL T-CELLS TOWARD IMMUNOPROTECTIVE  
CC THIS PHENOTYPE.  
CC  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC  
CC Name=1;  
CC  
CC IsoId=09JHJ8-1; Sequence=Displayed;  
CC Name=2; Synonym=B;  
CC  
CC IsoId=09JHJ8-2; Sequence=VSP\_002521;  
CC  
CC -1- TISSUE SPECIFICITY: ISOFORM 1 HIGHEST EXPRESSION IN LYMPHOID  
CC TISSUES, SUCH AS SPLEEN (MOSTLY IN THE MARGINAL ZONE), LYMPH NODES  
CC (PARTICULARLY IN THE CORTEX AND IN BOTH PRIMARY AND SECONDARY  
CC FOLLICLES), THYMUS (PREDOMINANTLY IN THE MEDULLA) AND PEYER'S  
CC PATCHES (MOSTLY IN THE FOLLICLES), LOWER LEVELS IN MANY

CC		NONMYELOID TISSUES, SUCH AS BRAIN, HEART, KIDNEY, LIVER, LUNG,
CC		SKELETAL MUSCLE AND TESTIS. PRESENT ON FRESHLY ISOLATED SPLENIC B-
CC		- CELLS, DENDRITIC CELLS AND MACROPHAGES. THE EXPRESSION OF
CC		ISOFORM 2 IS RESTRICTED TO HEART, SPLEEN AND KIDNEY.
CC		DEVELOPMENTAL STAGE: DETECTED EARLY IN HEMOPOIESIS; IN THE YOLK
CC		SAC AT 11.5 AND 12.5 DPC AND, TO A LESSER EXTENT, IN THE LAYER AT
CC		14.5 DPC.
CC		- SIMILARITY: Belongs to the immunoglobulin superfamily. BTN/MOG
CC		family.
CC		- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC		- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC		or send an email to license@isb-slb.ch).
CC		-----
DR		EMBL; AF216747; AAF45149.1; -
DR		EMBL; AF199027; AAF44738.1; -
DR		EMBL; AX100591; CAC36463.1; -
DR		EMBL; AX100593; CAC36464.1; -
DR		EMBL; AF394451; AAK77544.1; --
DR		MGI; MGI:1354701; ICob1.
DR		GO; GO:0005576; C:extracellular; IDA.
DR		GO; GO:0016021; C:integral to membrane; NAS.
DR		GO; GO:0003793; P:defense/immunity protein activity; NAS.
DR		GO; GO:0005102; F:receptor binding; TAS.
DR		GO; GO:0045190; P:isotype switching; NAS.
DR		GO; GO:0042104; P:positive regulation of activated T-cell pro. . ; TAS.
DR		GO; GO:0045404; P:positive regulation of interleukin-4 biosyn. . ; IMP.
DR		GO:GO:0007165; P:signal transduction; NAS.
DR		GO; GO:0042110; P:T-cell activation; NAS.
DR		InterPro; IPR007110; Ig-like.
DR		InterPro; IPR003599; IG.
DR		Pfam; PF00047; Ig; 1.
DR		SMART; SMO0409; IG; 1.
KM		PROSITE; PS50835; IG LIKE; 2.
KW		B-cell activation; Immune response; Glycoprotein;
KW		Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW		Alternative splicing.
FT	1	BY SIMILARITY.
FT	46	ICOS LIGAND.
FT	47	EXTRACELLULAR (POTENTIAL).
FT	277	POTENTIAL.
FT	298	CYTOPLASMIC (POTENTIAL).
FT	299	IG-LIKE V-TYPE.
FT	49	IG-LIKE C2-TYPE.
FT	167	POLY-LED.
FT	31	POLY-ALA.
FT	289	POTENTIAL.
FT	62	POTENTIAL.
FT	185	POTENTIAL.
FT	71	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	120	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	163	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	200	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	213	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	213	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	252	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	265	N-LINKED (GLCNAC. . ) (POTENTIAL).
FT	321	HA -> TWAPVQDYLIPIRLYMSFCLKIKRGIP (in isoform 2).
FT	322	/FTid=VSP_002521.
FT	237	R -> H (IN REF. 4 AND 5; CAC36464).
FT	322 AA;	SSCBRA4AD12E47B6 CRC64;
FT	SEQUENCE	
Query Match		7.4%; Score 97.5; DB 1; Length 322;
Best Local Similarity		21.5%; Pred. No. 2.2;
Matches		63; Conservative 39; Mismatches 104; Indels 87; Gaps 13;
Gy	10 LLVLVLGALLPAATQGKNRVYLGGKKGYVEIJCTASQRK----	SIOFHMKNSN-QIHK-- 61

Db 35 LLLLSLCAASAEVEGAVY-----GSNVVLSCIDPHRRHNLGLVYVWQIENPEVSVTY 90  
 QY 62 -----LGNQSGFLTKGSPKLANDRARRSLMDQGNFLIITKLIKEDSDTYICE--- 110  
 Db 91 YLPKSPGIVNDSSYKNGRHLSTL-----DSMK-----QGNFSLYKKNVLPDQTOEFTORVFM 142  
 QY 111 -----VEDQKEEYQVLVFGLTANSDPHLLOQGSFLTLESPPGSSPSVQCRPRG---K 161  
 Db 143 NTATELVKILKEEVRLVY---AANFSTPVISTD-----SSNFGQERTTYCKSNKNGYPER 194  
 QY 162 NIOGKTLVSQLELODGTWCTVLOKQKVEFKIDIVPRASALPAP----- 209  
 Db 195 NLVYINTTDSLID-----TALQNTVYINKGLYDVISTLRLEPMTSRGVDLCVE 245  
 QY 210 -----PTGSALPDPQTA--SALPDPAAALPALAVISFLL 244  
 Db 246 NVALHONITSISOAESFTGNNTKNPQETHNNELKVLVPLVALAAAFVSFI 298

RESULT 83  
 SMP\_COTUA ID SMP\_COTUA STANDARD; PRT; 620 AA.  
 AC Q92154;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Schwann cell myelin protein precursor (Siglec-4b).  
 GN SMP.  
 OS Coturnix coturnix japonica (Japanese quail).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Coturnix.  
 NC NCBI\_Taxid=93934;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-28; 120-132; 135-157 AND 563-571.  
 RX MEDLINE=92153423; PubMed=1739462;  
 RA Dulac C., Tropak M.B., Cameron-Curry P., Rossier J., Marshak D.R., Roder J., le Donnat N.M.;  
 RT "Molecular characterization of the Schwann cell myelin protein, SMP: structural similarities within the immunoglobulin superfamily.";  
 RL Neuron 8:323-334 (1992).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Exclusively expressed by myelinating and nonmyelinating schwann cells and oligodendrocytes.  
 CC -1- DEVELOPMENTAL STAGE: First synthesized at embryonic day 5, it remains expressed by cultured Schwann cells.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC (sialic acid binding Ig-like lectin) family.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; S83711; AAB21466.1; -  
 CC PIR; JH0593; JH0593.  
 DR HSSP; P56276; ITLK.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR Pfam; PF00047; Ig\_2.  
 DR SMART; SM00408; IgC2; 2.  
 DR PROSITE; PSS0835; IG\_LIKE; 2.  
 KW Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;  
 KW Immunoglobulin domain; Repeat.  
 FT SIGNAL 1 17  
 FT CHAIN 18 620 SCHWANN CELL MYELIN PROTEIN.  
 FT DOMAIN 18 516 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 517 536 POTENTIAL.  
 FT DOMAIN 537 620 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 28 106 IG-LIKE V-TYPE.  
 FT DOMAIN 151 233 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 239 322 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 325 407 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 414 495 IG-LIKE C2-TYPE 4.  
 FT DISULFID 35 164 BY SIMILARITY.  
 FT DISULFID 40 99 BY SIMILARITY.  
 FT DISULFID 158 216 BY SIMILARITY.  
 FT DISULFID 260 304 BY SIMILARITY.  
 FT DISULFID 346 391 BY SIMILARITY.  
 FT DISULFID 420 429 BY SIMILARITY.  
 FT DISULFID 431 488 BY SIMILARITY.  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 620 AA; 66943 MW; 004B3E7EDC18FBA CRC64;

Query Match 7.4%; Score 97.5; DB 1; Length 620;  
 Best Local Similarity 20.3%; Pred. No. 5;  
 Matches 48; Conservative 36; Mismatches 91; Indels 61; Gaps 10;

QY 33 KEDTVELTCTAQSOKSIQFHMKNSNOIKILNGSGFLTKGSPKLANDRARRSLMDQGNF 92  
 Db 252 EGSVDVLCGEARGPAPLISWFRGSEV-----LREPGK-----NL 287  
 QY 93 PLTIKLIKIEDSDTYICEVEDQ-----KEVQVLVFG-----GLTANSDPHLLOQ--SLTVLT 142  
 Db 288 RLILSVGPDGDSFECVAVENRHRNRLQLRVAVAPAPVINGSLWVSGDPVSVTR 347  
 QY 143 LESPPGSSPSVQCRSEPRGNIOGK-----TISVQLELODGTWCTVLOQ-- 190  
 Db 348 ASESPPAIIITV-----LRGKVMAAIYEDHTWEMRPARPEDGGTYSC-VAENQH 397  
 QY 191 --KXVEFKIDIVPRASALPAP-----PTGSALPDPQTA--SALPDPAAALPALAVIS 241  
 Db 398 GASSTSFNISVEYPLVLPASRCTAGGAGVRCVCMVNSIPDSSLVPELPTRNQTVS 453

RESULT 84  
 PIGR\_MOUSE ID PIGR\_MOUSE STANDARD; PRT; 771 AA.  
 AC Q70570;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE PolymERIC-Immunoglobulin receptor precursor (Poly-Ig receptor) (PIGR) [contains: Secretory component].  
 GN PIGR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 NC NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
 RX MEDLINE=95138517; PubMed=7836758;  
 RA Piskurich J.F., Blanchard M.H., Youngman K.R., France J.A., Kaetzel C.S.;  
 RT "Molecular cloning of the mouse polymERIC Ig receptor. Functional regions of the molecule are conserved among five mammalian species.";  
 RL J. Immunol. 154:1735-1747 (1995).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SvJ;  
 RX MEDLINE=98072444; PubMed=9409786;  
 RA Martin M.G., Gutierrez E.M., Lam J.T., Li T.W.H., Wang J.;  
 RT "Genomic cloning and structural analysis of the murine polymERIC receptor (PIGR) gene and promoter region.";  
 RL Gene 201:189-197 (1997).

```

RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129; TISSUE=Liver;
RA      MEDLINE=99410926; PubMed=10481312;
RX      de Groot N., van Kuik-Romeijn P., Lee S.H., de Boer H.A.;
RT      "Over-expression of the murine polymeric immunoglobulin receptor gene
RL      in the mammary gland of transgenic mice.";
       Transgenic Res. 8:125-135(1999).
CC      -I- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC      BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC      TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC      DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC      EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC      TRANSMEMBRANE SEGMENT (BY SIMILARITY).
CC      -I- SUBCELLULAR LOCATION: Type I membrane protein. Also secreted.
CC      -I- SIMILARITY: Contains 5 Immunoglobulin-like V-type domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration between
CC      the Swiss Institute of Bioinformatics and the EMBL outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to licenses@isb-sdb.ch).
CC      -----
DR      EMBL; U06431; AA67440.1; -.
DR      EMBL; U83434; AAC53585.1; -.
DR      EMBL; U83427; AAC53585.1; JOINED.
DR      EMBL; U83428; AAC53585.1; JOINED.
DR      EMBL; U83429; AAC53585.1; JOINED.
DR      EMBL; U83430; AAC53585.1; JOINED.
DR      EMBL; U83431; AAC53585.1; JOINED.
DR      EMBL; U83432; AAC53585.1; JOINED.
DR      EMBL; U83433; AAC53585.1; JOINED.
DR      EMBL; Y16524; CAA76272.1; -.
DR      EMBL; Y16525; CAA76272.1; JOINED.
DR      EMBL; Y16526; CAA76272.1; JOINED.
DR      EMBL; Y16527; CAA76272.1; JOINED.
DR      EMBL; Y16528; CAA76272.1; JOINED.
DR      EMBL; Y16529; CAA76272.1; JOINED.
DR      EMBL; Y16530; CAA76272.1; JOINED.
DR      EMBL; Y16531; CAA76272.1; JOINED.
DR      EMBL; Y16532; CAA76272.1; JOINED.
DR      WGI; MGI:103080; P1gr.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003599; Ig.
DR      Pfam; PF00047; Ig_5.
DR      SMART; SMO0409; IG_3.
DR      PROSITE; PS50835; IG_LIKE; 3.
RW      KMW
KW      Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal.
FT      SIGNAL              1          18
FT      CHAIN              19          771
FT      CHAIN              19          771
FT      DOMAIN            19          645
FT      TRANSMEM         646          668
FT      DOMAIN            669          771
FT      DOMAIN            21          120
FT      DOMAIN            135          237
FT      DOMAIN            245          351
FT      DOMAIN            352          457
FT      DOMAIN            463          563
FT      DISULFID           40          110
FT      DISULFID           152          220
FT      DISULFID           257          324
FT      DISULFID           370          440
FT      DISULFID           484          546
FT      CARBOHYD           90
FT      CARBOHYD           147          147
FT      CARBOHYD           170          170
FT      CARBOHYD           206          206
FT      CARBOHYD           420          420
FT      CARBOHYD           471          471
FT      CONFLICT           159          159

```

FT	CONFLICT	396	396	V -> A (IN REF. 1)
FT	CONFLICT	620	620	G -> R (IN REF. 1)
SQL	SEQUENCE	771 AA;	84998 MW;	78C81302BC710730 CRC64;
	Query Match		7.4%;	Score 97.5; DB 1; Length 771;
	Best Local Similarity		23.0%;	Pred. No. 6.5;
	Matches	59;	Conservative 43;	Mismatches 91; Indels 63; Gaps 13;
QY	15 QALPPATQGNK--VVLGKGGTVEVLTCTASQKKSIOF---- <td>65</td> <td></td> <td></td>	65		
DB	342 QLPVNESTTIPNRRSVVGVGTGGVLAICAPNPRESSSLKWCWREGDGNCHCEVLVGTQ	401		
QY	66 GSFLLTKGSPKINDADADRRSLMDQ---GNPFLIKLIKIEDSDYIC-----	109		
DB	402 -----AQOEBEYEGLLAFDPGNGTGYVILNQLTTEBAGFYWCLCTNGDSRWRTTIE	453		
QY	110 -EVEDQKEEVLAFGLGTANSDTHLQGSQTLTLIESPPG--SSPSVOCR-SPRGKNI--	163		
DB	454 LQVNEAIRREPLVEY---TPQNAITAVL-GETFTVSCHPCKRYSEKWKCKSNKGCHILP	509		
QY	164 ---QSGKTLVS-----QLELDQSGTWCTTVLONQKKEVFKDIDVRASALP	207		
DB	510 SHDEGARGSSVSCQSSQSLVSMTLNIPVSKEDEGWYGVQGGQTYGETTAIYI----AVE	565		
QY	208 APPTGSALPPDQTASA 223			
DB	566 ERTGSSSHVNPPTDANA 581			
	RESULT 85			
	VER1_MOUSE	STANDARD;	PRT;	1333 AA.
ID	VER1_MOUSE			
AC	P35969;	055094;	Q61517;	
DT	01-JUN-1994	(Rel. 29,	Created)	
DT	01-JUN-1994	(Rel. 29,	Last sequence update)	
DT	10-OCT-2003	(Rel. 42,	Last annotation update)	
DE	Vascular endothelial growth factor receptor 1 precursor (EC 2.7.1.112)			
DE	(VEGFR-1) (Tyrosine-protein kinase receptor FLT) (Flt-1) (Embryonic			
GN	receptor kinase 2).			
OS	Flt1 OR FLT OR EMRK2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutehia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Balb/c; TISSUE=Neonatal brain, and Placenta;			
RX	FMINLINE=93330572; PubMed=8393164;			
RX	Fminely H., Kelleher K., Morris G.E., Bean K., Werberg D.M.,			
RA	Kritz R., Morris J.C., Sookdeo H., Turner K.J., Wood C.R.;			
RT	"Molecular cloning of murine FLT and FLT4.";			
RL	Oncogene 8:2293-2298 (1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	FMINLINE=94181281; PubMed=8134130;			
RA	Choi K., Wall C., Hanratty R., Keller G.;			
RT	"Isolation of a gene encoding a novel receptor tyrosine kinase from			
RL	differentiated embryonic stem cells.";			
RL	Oncogene 9:1261-1266 (1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Lung;			
RX	FMINLINE=98201625; PubMed=9524283;			
RA	Kondo K., Hiratsuka S., Subbalakshmi E., Matsushine H., Shibuya M.;			
RT	"Genomic organization of the flt-1 gene encoding for vascular			
RT	endothelial growth factor (VEGF) receptor-1 suggests an intimate			
RL	evolutionary relationship between the 7-ig and the 5-ig tyrosine			
RL	kinase receptors.";			
RL	Gene 208:297-305 (1998).			
CC	-1- FUNCTION: RECEPTOR FOR VEGF AND PLGF. HAS A TYROSINE-PROTEIN			
CC	KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM			
CC	PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF			

```

CC VASCULAR PERMEABILITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC CSF-1/PDGF receptor subfamily.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; L07297; AAA40078.1; -
DR EMBL; X78568; CAA55311.1; -
DR EMBL; D88689; BAA24498.1; -
DR PIR; I78875; I78875.
DR PIR; S49010; S49010.
DR HSSP; P17948; IQSV.
DR MGD; MGI.95558; PLC1.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig C2.
DR InterPro: IPR00719; Prot_kinase.
DR InterPro: IPR001824; ReceptTyKinsIII.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR008266; Tyr_kinase_AS.
DR Pfam; PF00047; Ig; 6.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS50835; IG_LIKE; 5.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
KW Angiogenesis; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW ATP-binding; Receptor; Transmembrane; Signal; Immunoglobulin domain;
KW Repeat; Glycoprotein.
FT CHAIN 1 22
FT SIGNAL 23 1333
FT TRANSMEM 23 759
FT TRANSMEM 760 781
FT DOMAIN 782 1333
FT DOMAIN 124 124
FT DOMAIN 152 215
FT DOMAIN 231 328
FT DOMAIN 334 429
FT DOMAIN 430 550
FT DOMAIN 557 656
FT DOMAIN 662 748
FT DOMAIN 828 1158
FT NP_BIND 834 842
FT BINDING 862 862
FT ACT_SITE 1022 1022
FT MOD_RES 1053 1053
FT DISULFID 53 108
FT DISULFID 159 208
FT DISULFID 253 312
FT DISULFID 455 536
FT DISULFID 578 637
FT DISULFID 683 732
FT CARBOHYD 101 101
FT CARBOHYD 165 165
FT CARBOHYD 197 197
FT CARBOHYD 252 252
FT CARBOHYD 324 324
FT CARBOHYD 418 418
FT CARBOHYD 475 475

```

```

FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 MISSING (IN REF. 2).
FT CONFICT 211 211 MISSING (IN REF. 2).
FT CONFICT 245 245 H -> L (IN REF. 2).
FT CONFICT 603 603 H -> N (IN REF. 2).
FT CONFICT 609 615 KMTTDD -> NGHSSS (IN REF. 2).
FT CONFICT 696 696 F -> L (IN REF. 2).
FT CONFICT 734 734 A -> S (IN REF. 2).
FT CONFICT 765 765 C -> Y (IN REF. 2).
FT CONFICT 820 820 K -> N (IN REF. 2).
FT CONFICT 1009 1009 G -> R (IN REF. 2).
FT CONFICT 1181 1181 S -> G (IN REF. 3).
FT CONFICT 1181 1181 S -> N (IN REF. 2).
FT CONFICT 1193 1194 LP -> RG (IN REF. 2).
FT CONFICT 1278 1279 KS -> PR (IN REF. 2).
SQ SEQUENCE 1333 AA; 149875 MM; C06533B7ECBC404C CRC64;

Query Match 7.4%; Score 97.5; DB 1; Length 1333;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 45; Conservative 30; Mismatches 71; Indels 43; Gaps 9;

QY 33 KEDTVELCTGASQKSIQFHMKNNSQIKILNGSGFLTGSPSKLNRADSRRLMPQGNF 92
DB 570 EGGDLKSLCVNFFLYRDITW---ILKRTVNNRTMHSISKQKMTTQDYSTL----- 620
QY 93 PLIKNKLIKEDSDPYICEV-----EDQKEVQLVFGLTANSDFHLOGQSILTLLESP 146
DB 621 NLVTKNKVSLDSEGTVCGRANITYTGEDILRKNEVLVRDSEA---PHLLQNLSS---DYEVS 674
QY 147 PGSSPEVQCRSPRG-----KN---IQ-----GKTLVSQLELDQSGTWCT 185
DB 675 ISGSLTLDQA-RGVAPQITWFKNNHKIQDEPGIILGSGNSTLFLERVTEDEGVYRCR 733
QY 186 VLQNKQKVE 194
DB 734 ATNQKQAVE 742

RESULT 86
UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362.
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR CO9D.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Briscot N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains";
RL V. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```

CC	- FUNCTION: Structural component of the muscle M-line. Myofibrament
CC	lattice assembly begins with positional cues laid down in the
CC	basement membrane and muscle cell membrane. UNC-89 responds to
CC	these signals, localizes, and then participates in assembling an
CC	M-line.
CC	- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC	- SIMILARITY: Contains 1 DBI-homology (DB) domain.
CC	- SIMILARITY: Contains 1 fibronectin type III domain.
CC	- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC	- SIMILARITY: Contains 1 PH domain.
CC	- SIMILARITY: Contains 5 RCDSD domains.
CC	- SIMILARITY: Contains 1 SH3 domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as their content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).
CC	-----
DR	EMBL; U33058; AAB00542.1; -.
DR	EMBL; AF003131; AAB54132.2; -.
DR	PDB; 1PHO: 20-DEC-00.
DR	WormRep; C09D1.1; CE30426.
DR	InterPro; IPR008957; FN_III-like.
DR	InterPro; IPR003961; FN_III.
DR	InterPro; IPR007110; IG-1-like.
DR	InterPro; IPR003598; IG_C2.
DR	InterPro; IPR003066; IG_MHC.
DR	InterPro; IPR001849; RH.
DR	InterPro; IPR007850; RCDSD.
DR	InterPro; IPR00219; RhogGEF.
DR	InterPro; IPR001452; SH3.
DR	Pfam; PF000441; fn3; 1.
DR	Pfam; PF00047; Ig; 47.
DR	Pfam; PF00169; PH; 1.
DR	Pfam; PF05177; RCDSD; 5.
DR	Pfam; PF00621; RhogGEF; 1.
DR	Pfam; PF000148; SH3; 1.
DR	SMART; SMO0408; IGC2; 23.
DR	SMART; SMO0325; RhogGEF; 1.
DR	SMART; SMO0326; SH3; 1.
DR	PROSITE; PSS0010; DH_2; 1.
DR	PROSITE; PSS0835; IG_LIKE; 49.
DR	PROSITE; PSS0003; PH DOMAIN; 1.
DR	PROSITE; PSS0002; SH3; 1.
KW	Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW	3D-structure.
FT	DOMAIN 63 127 SH3.
FT	DOMAIN 152 330 DH.
FT	DOMAIN 342 498 PH.
FT	DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT	DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT	DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT	DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT	DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT	DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT	DOMAIN 1272 1315 THR-RICH.
FT	DOMAIN 1375 1475 RCDSD 1.
FT	DOMAIN 1479 1585 RCDSD 2.
FT	DOMAIN 1597 1695 RCDSD 3.
FT	DOMAIN 1700 1799 RCDSD 4.
FT	DOMAIN 1800 1860 RCDSD 5.
FT	DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT	DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT	DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT	DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT	DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT	DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT	DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT	DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT	DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.

Query Match	7.4%	Score 97.5	DB 1	Length 6632
Best Local Similarity	24.4%	Pred. No. 92		
Matches	50	Conservative	28	Mismatches 60; Indels 67; Gaps 10
27	KVLGKGGIVNELCTASQKSIQFHWKNSNLIKILGNQSSPLTKGPSKLNDRAISR	85		
FT	DOMAIN	2987	2980	IG-LIKE C2-TYPE 16
FT	DOMAIN	2994	3081	IG-LIKE C2-TYPE 17
FT	DOMAIN	3087	3183	IG-LIKE C2-TYPE 18
FT	DOMAIN	3189	3280	IG-LIKE C2-TYPE 19
FT	DOMAIN	3286	3376	IG-LIKE C2-TYPE 20
FT	DOMAIN	3384	3469	IG-LIKE C2-TYPE 21
FT	DOMAIN	3482	3572	IG-LIKE C2-TYPE 22
FT	DOMAIN	3580	3667	IG-LIKE C2-TYPE 23
FT	DOMAIN	3686	3777	IG-LIKE C2-TYPE 24
FT	DOMAIN	3920	4009	IG-LIKE C2-TYPE 25
FT	DOMAIN	4018	4106	IG-LIKE C2-TYPE 26
FT	DOMAIN	4109	4201	IG-LIKE C2-TYPE 27
FT	DOMAIN	4212	4297	IG-LIKE C2-TYPE 28
FT	DOMAIN	4302	4387	IG-LIKE C2-TYPE 29
FT	DOMAIN	4400	4485	IG-LIKE C2-TYPE 30
FT	DOMAIN	4489	4580	IG-LIKE C2-TYPE 31
FT	DOMAIN	4588	4678	IG-LIKE C2-TYPE 32
FT	DOMAIN	4681	4771	IG-LIKE C2-TYPE 33
FT	DOMAIN	4873	4961	IG-LIKE C2-TYPE 34
FT	DOMAIN	4965	5057	IG-LIKE C2-TYPE 35
FT	DOMAIN	5067	5160	IG-LIKE C2-TYPE 36
FT	DOMAIN	5171	5260	IG-LIKE C2-TYPE 37
FT	DOMAIN	5277	5366	IG-LIKE C2-TYPE 38
FT	DOMAIN	5383	5472	IG-LIKE C2-TYPE 39
FT	DOMAIN	5487	5578	IG-LIKE C2-TYPE 40
FT	DOMAIN	5595	5685	IG-LIKE C2-TYPE 41
FT	DOMAIN	5701	5790	IG-LIKE C2-TYPE 42
FT	DOMAIN	5815	5904	IG-LIKE C2-TYPE 43
FT	DOMAIN	5925	6014	IG-LIKE C2-TYPE 44
FT	DOMAIN	6038	6130	IG-LIKE C2-TYPE 45
FT	DOMAIN	6275	6368	IG-LIKE C2-TYPE 46
FT	DOMAIN	6413	6502	IG-LIKE C2-TYPE 47
FT	DOMAIN	6507	6596	IG-LIKE C2-TYPE 48
FT	DOMAIN	6507	6596	IG-LIKE C2-TYPE 49
FT	DISULFID	568	621	POTENTIAL.
FT	DISULFID	2908	2975	POTENTIAL.
FT	DISULFID	3015	3065	POTENTIAL.
FT	DISULFID	3707	3759	POTENTIAL.
FT	DISULFID	3826	3890	POTENTIAL.
FT	DISULFID	5092	5157	POTENTIAL.
FT	DISULFID	5288	5350	POTENTIAL.
FT	DISULFID	5508	5560	POTENTIAL.
FT	DISULFID	5616	5669	POTENTIAL.
FT	DISULFID	5722	5764	POTENTIAL.
FT	DISULFID	5836	5901	POTENTIAL.
FT	DISULFID	5946	5998	POTENTIAL.
FT	DISULFID	6036	6171	POTENTIAL.
FT	DISULFID	6421	6486	POTENTIAL.
FT	CONFLICT	2137	2137	AKA -> P (IN REF. 1).
FT	CONFLICT	2245	2247	A -> P (IN REF. 1).
FT	CONFLICT	2258	2258	E -> G (IN REF. 1).
FT	CONFLICT	2284	2284	M -> I (IN REF. 1).
FT	CONFLICT	2297	2297	A -> G (IN REF. 1).
FT	CONFLICT	3531	3531	DAGEY -> RRRRI (IN REF. 1).
FT	CONFLICT	3884	3888	A -> V (IN REF. 1).
FT	CONFLICT	3929	3929	A -> P (IN REF. 1).
FT	CONFLICT	5134	5134	T -> S (IN REF. 1).
FT	CONFLICT	5145	5145	G -> A (IN REF. 1).
FT	CONFLICT	5185	5185	K -> N (IN REF. 1).
FT	CONFLICT	5199	5199	L -> F (IN REF. 1).
FT	CONFLICT	5202	5202	F -> L (IN REF. 1).
FT	CONFLICT	5213	5213	A -> G (IN REF. 1).
FT	CONFLICT	6178	6178	K -> E (IN REF. 1).
FT	CONFLICT	6268	6268	
SO	SEQUENCE	6632	AA; 731665	NW; 262D3EDD62960E89 CRC64;





FT	CARBOHYD	478	478		(POTENTIAL).
FT	VARSPLIC	702	726		STPTSGGLGAAVGLIVLIFVLL -> TLGSPSTSSFV
FT					SLLSVTLVLIC (in isoform D).
FT					/Frid=VSP 002583.
FT	VARSPLIC	727	1091		Missing (in isoform D).
FT					/Frid=VSP .002584.
FT	VARSPLIC	771	809		Missing (in isoform B).
FT					/Frid=VSP .002585.
FT	VARSPLIC	810	1070		Missing (in isoform C).
FT					/Frid=VSP .002586.
SO	SEQUENCE	1091 AA,	117415 MW;	B151567002D808ED	CNC64;
Query Match		7.4%;	Score 97;	DB 1;	Length 1091;
Best Local Similarity		21.4%;	Pred. No. 11;		
Matches		51;	Conservative	27;	Mismatches 84; Indels 76; Gaps 10;
OY	35 DTVELTCTASQKKSIOFHMKN-----SNCKILGNOSGFLTGKPSKLNDRADRSRLMDQ	89			
Db	323 DQITLTEASGDPIPTITWTSTRNISNEKITL--DRIYVRSHAVYS-----	369			
OY	90 GNFPILIKNLKIEDSDTYICEVEDOKEEVQLLVFGLTANSDTHL-----LOGSL	139			
Db	370 ----LFLKEIQTVDAGEYVCTAAN-----TIQDSQAMYLEVQAPPLQGPA	413			
OY	140 TLTIESPPGSSPVQCR-----SPRKNIQGGKTLISVSOLE----	176			
Db	414 VYTWE---GNQVNITCEVFAYPSAVISWFRDGLLPBSNSYNKIKNTPASATLEVTPTS	470			
OY	177 -ODSGTWTCVTLQ--NOKKVEFKIDIVPRASALPAPTGSALPDPOFASALPDPFAAS	231			
Db	471 ENDFGNNTCTIAVMRIGESSEF---ILVQADPTSPSIDRVEEYSSARARVPDEPEART	525			
RESULT 89					
DCC_HUMAN	ID_DCC_HUMAN	STANDARD;	PRT;	1447	AA.
AC	P34146;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).				
GN	DCC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=95011532; PubMed=7926722;				
RA	Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,				
RA	Vogelstein B.;				
RT	"The DCC gene product in cellular differentiation and colorectal				
RT	tumorigenesis."				
RL	Genes Dev. 8:1174-1183(1994).				
RN	[2]				
RN	SEQUENCE OF 1-750 FROM N.A.				
RX	MEDLINE=90100559; PubMed=2294591;				
RA	Fearon E.R., Cho K.R., Niigro J.M., Kern S.E., Simons J.W.,				
RA	Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,				
RT	Vogelstein B.;				
RT	"Identification of a chromosome 18q gene that is altered in				
RT	colorectal cancers."				
RL	Science 247:49-56(1990).				
RN	[3]				
RN	SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).				
RX	MEDLINE=9112517; PubMed=1991322;				
RA	Niigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,				
RA	Oliver J.D., Kinzler K.W., Vogelstein B.;				
RT	"Scrambled exons."				
RL	Cell 64:607-613(1991).				
RN	[4]				
RN	GENE STRUCTURE, AND VARIANT CARCINOMA HIS-1375.				
RX	MEDLINE=94245241; PubMed=8186295;				



RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,  
 RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;  
 RT "The DCC gene: structural analysis and mutations in colorectal  
 RT carcinomas";  
 RN Genomics 19:525-531(1994).  
 RN (5)  
 RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.  
 RX MEDLINE=94243823; PubMed=8187090;  
 RA Miyake S., Nagai K., Yoshino K., Ota M., Endo M., Yuasa Y.;  
 RT "point mutations and allelic deletion of tumor suppressor gene DCC in  
 RT human esophageal squamous cell carcinomas and their relation to  
 RT metastasis";  
 RL Cancer Res. 54:3007-3010(1994).  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL  
 CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.  
 CC -1- DISEASE: Colorectal tumors that lost their capacity to  
 CC differentiate into mucus producing cells uniformly lack DCC  
 CC expression. Inactivation of DCC due to allelic deletion and/or  
 CC point mutations may cause both lymphatic and hematogenous  
 CC metastasis of esophageal squamous cell carcinomas.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.  
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 6 fibronectin type III domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC  
 DR EMBL: X76132; CAAS3735.1; -  
 DR EMBL: M32292; AAA52751.1; -  
 DR EMBL: M32286; AAA52174.1; -  
 DR EMBL: M32288; AAA52175.1; ALT\_SEQ.  
 DR EMBL: M32290; AAA52176.1; -  
 DR EMBL: M63696; AAA52177.1; -  
 DR EMBL: M63700; AAA52178.1; -  
 DR EMBL: M63702; AAA52179.1; -  
 DR EMBL: M63718; AAA52180.1; -  
 DR EMBL: M63698; AAA52181.1; -  
 DR PIR: A54100; A54100.  
 DR HSSP: P56276; ITLK.  
 DR Genew; HGNC:2701; DCC.  
 DR MIM; 120470; -  
 DR GO; GO:0004888; F:transmembrane receptor activity; TMS.  
 DR GO; GO:0007409; P:axonogenesis; TMS.  
 DR GO; GO:0006917; P:induction of apoptosis; TMS.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003962; FNIII subd.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003598; IG\_C2.  
 DR Pfam; PF00041; fn3; 6.  
 DR Pfam; PF00047; Ig; 4.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR SMART; SMO0060; FN3; 6.  
 DR SMART; SMO0408; IGC2; 3.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
 KW Repeat; Anti-oncogene; Disease mutation; Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447  
 FT DOMAIN 26 1097 TUMOR SUPPRESSOR PROTEIN DCC.  
 FT TRANSMEM 1098 1122 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 1123 1447 POTENTIAL.  
 FT DOMAIN 26 135 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 139 229 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 234 326 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 331 416 IG-LIKE C2-TYPE 3.  
 FT IG-LIKE C2-TYPE 4.

FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.  
 FT DISULFID 61 117 BY SIMILARITY.  
 FT DISULFID 161 212 BY SIMILARITY.  
 FT DISULFID 261 310 BY SIMILARITY.  
 FT DISULFID 352 400 BY SIMILARITY.  
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 168 168 M -> T (in esophageal carcinoma).  
 FT VARIANT 168 168 /FtId=VAR\_003909.  
 FT VARIANT 201 201 R -> G.  
 FT VARIANT 201 201 /FtId=VAR\_003910.  
 FT VARIANT 1375 1375 P -> H (in a colorectal carcinoma).  
 FT CONFLICT 138 138 /FtId=VAR\_003911.  
 FT CONFLICT 233 329 MISSING (IN REF. 3).  
 FT CONFLICT 421 421 MISSING (IN REF. 3).  
 SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;

Query Match 7.4%; Score 97; DB 1; Length 1447;  
 Best Local Similarity 23.7%; Pred. No. 15;  
 Matches 45; Conservative 26; Mismatches 77; Indels 42; Gaps 8;

OY 34 GDTVELCTASQKRSIQFMKNSNCKILNQGSLFKRGPKKNDNRDRSRSLDQGNFP 93  
 DB 154 GDTVLLKCEVIGBEMPTTHQKNQ-----DLTPIP-----GDSRVVLPSC--A 196  
 OY 94 LIINKLKIEDSDTYICEVED-----QKEVQL-----VFGLTANSDFTLQSGS 138  
 DB 197 LQISRLLPGDGLRCARNPASRTGNEAVRLSPGLHRLQYLFQRPNNVAIRGKD 256  
 OY 139 LTL---TLSEPPGS-----SPVQCRSPKRNKIQGSKTSLVSQLELQDSGTWTCTYKLN 189  
 DB 257 AVLECCVSGVPPFPFTWLRGSEVQLRSKK-YSLGGSNLLISVTTDDSGMYTCVITYK 315  
 OY 190 QKTYEKKIDI 199  
 DB 316 NENISASAEI 325

RESULT 90  
 KY1J HUMAN  
 ID KY1J HUMAN STANDARD; PRT; 117 AA.  
 AC P01602;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE IG kappa chain V-I region HK102 precursor (Fragment).  
 GN IGKV1-5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=81098966; PubMed=6779204;  
 RA Bentley D.L., Rabbits T.H.;  
 RT "Human immunoglobulin variable region genes -- DNA sequences of two V  
 RT kappa genes and a pseudogene";  
 RL Nature 288:730-733(1980).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; J00245; AAA59087.1; -  
 CC EMBL; J00001; CAA77292.1; -  
 CC PIR; A01882; KJH12.  
 CC HSSP; P01607; IRE1.  
 CC Genew; HGNC; 5741; IGKV1-5.  
 CC GO; GO:0005576; C:extracellular; NAS.  
 CC GO; GO:0003823; F:antigen binding; NAS.  
 CC GO; GO:0006955; P:immune response; NAS.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003596; Ig\_v.  
 CC Pfam; PF00047; Ig; 1.  
 CC SMART; SM00406; IGV; 1.  
 CC PROSITE; PS50835; IG LIKE; 1.  
 CC Immunoglobulin V region; Signal.  
 CC SIGNAL 1 22  
 FT CHAIN 23 >117 IG KAPPA CHAIN V-I REGION HK102.  
 FT DOMAIN 23 45 FRAMEWORK-1.  
 FT DOMAIN 46 56 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 57 71 FRAMEWORK-2.  
 FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 79 110 FRAMEWORK-3.  
 FT DOMAIN 111 >117 COMPLEMENTARITY-DETERMINING-3.  
 FT DISULFID 45 110 BY SIMILARITY.  
 FT NON TER 117  
 SQ SEQUENCE 117 AA; 12768 MW; AD1DF3A40AF1A9B CRC64;

Query Match 7.3%; Score 96.5; DB 1; Length 117;  
 Best Local Similarity 31.7%; Pred. No. 0.77; Indels 19; Gaps 5;  
 Matches 38; Conservative 11; Mismatches 52;

QY 1 MNRGVPRLHLLVQLALPAA-----TQGNKVVLGKGDVVELTCTASOKSIOFHW- 53  
 Db 1 MDKRVPAQ--LIGILLMLPGAKCIDMTQSPSLTSVSDRYITTCRASGISSWLAWY 58

QY 54 --KNSNOIKILGNQGFLLTG-PSKLNDRADSRRLSDQGNFPIITKNLKIEDSPYICE 110  
 Db 59 QQKPKAKPLIYDASSLGSGVPSRFGSGSGT-----EFTLLTSLQPDFFATYQC 111

RESULT 91  
 ID PIGR RAT STANDARD; PRT; 769 AA.  
 AC P15083;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE [Contains: Immunoglobulin receptor precursor (Poly-Ig receptor) (PIGR) (PIGR).  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RX MEDLINE=6937826; PubMed=2776682;  
 RA Banting G., Brake B., Braghetta P., Luzzo J.P., Stanley K.K.;  
 RT "Intracellular targeting signals of polymeric immunoglobulin receptors are highly conserved between species."  
 RL FEBS Lett. 254:177-183(1989).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE. DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Also secreted.  
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like V-type domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; J15741; CAA33758.1; -  
 CC PIR; S05407; QRRGS.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003599; Ig.  
 CC Pfam; PF00047; Ig; 5.  
 CC SMART; SM00409; IG; 5.  
 CC PROSITE; PS50835; IG LIKE; 2.  
 CC Immunoglobulin domain; Repeat; Transmembrane; glycoprotein; Signal.  
 CC SIGNAL 1 18  
 FT CHAIN 19 769 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
 FT CHAIN 19 611 SECRETORY COMPONENT.  
 FT DOMAIN 19 643 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 644 666 POTENTIAL.  
 FT DOMAIN 667 769 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 21 126 IG-LIKE V-TYPE 1.  
 FT DOMAIN 135 237 IG-LIKE V-TYPE 2.  
 FT DOMAIN 240 341 IG-LIKE V-TYPE 3.  
 FT DOMAIN 353 457 IG-LIKE V-TYPE 4.  
 FT DOMAIN 463 563 IG-LIKE V-TYPE 5.  
 FT DISULFID 40 110 POTENTIAL.  
 FT DISULFID 152 220 POTENTIAL.  
 FT DISULFID 257 324 POTENTIAL.  
 FT DISULFID 370 440 POTENTIAL.  
 FT DISULFID 484 546 POTENTIAL.  
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 769 AA; 84798 MW; 5F849303400255A7 CRC64;

Query Match 7.3%; Score 96.5; DB 1; Length 769;  
 Best Local Similarity 21.9%; Pred. No. 7.8;  
 Matches 59; Conservative 46; Mismatches 107; Indels 57; Gaps 13;

QY 11 LVVLQALPAAATQGNKVVLGKGDVVELTCTASOKSIOF-----HW-----KNSNQIKILG 63  
 Db 343 LFVNEESTIP---NSRSVKGVTGSAIVCPYNPSESSSLKWKWEADENRCPLVG 399

QY 64 NQGSFLTKGPKLNDRADSRRLSDQ--GNPFLIKNLIKIEDSPYICEVEDQ--KEE 117  
 Db 400 TQ-----ALVQEGYEGRLALFDPGSGATVILNQLTQDSGYWCLTDGDSMRRT 451

QY 118 VOLLVFGLTANSDTHL-----LOGQSILTLTESPPG--SSPSVOCR-SPRKNI---- 163  
 Db 452 IELQVAEARKPKDLEVTLPONATAVIGETFTISCHYCKEYSGEKYCKWNSNGCHILPSH 511

QY 164 -QGKTLSTVS-----QLELDQSGMTCTVNLQNKVKEKIDIVPAASALPAP 209  
 Db 512 DEGARQSVSCDQSSQIVSMNTLNPVKKEDEGMWCGVKGQYVGETTAIV---AVERR 567

QY 210 PTGSALPPDPOTASA-LPDPASALPAL 237  
 Db 568 TRGSPHINPTDANAKADAPBEEMESSV 596

RESULT 92  
 ID SN MOUSE STANDARD; PRT; 1694 AA.  
 AC Q62230; O55216; Q62228; Q62229;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Staladhesin precursor (Stalac acid binding Ig-like lectin-1) (Siglec-1) (sheep erythrocyte receptor) (SER).

GN SN OR SA.  
 OC Mus musculus (Mouse).  
 OC Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=C57BL/6; TISSUE=Macrophage;  
 RX MEDLINE=95009950; PubMed=7925291;  
 RA Crocker P.R., Mucklow S., Boukxon V., McWilliam A., Willis A.C.,  
 RA Gordon S., Milon G., Kelm S., Bradfield P.;  
 RT "Sialoadhesin, a macrophage sialic acid binding receptor for  
 RT haemopoietic cells with 17 immunoglobulin-like domains";  
 RL EMO J. 13:4490-4503 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96051930; PubMed=9383289;  
 RA Mucklow S., Gordon S., Crocker P.R.;  
 RT "Characterization of the mouse sialoadhesin gene, Sn";  
 RL Mamm. Genome 8:934-937 (1997).  
 RN [3]  
 RP CHARACTERIZATION  
 RC STRAIN=C57BL/6; TISSUE=Spleen;  
 RX MEDLINE=91266893; PubMed=2050106;  
 RA Crocker P.R.;  
 RT "Purification and properties of sialoadhesin, a sialic acid-binding  
 RT receptor of murine tissue macrophages";  
 RL EMO J. 10:1661-1669 (1991).  
 RN [4]  
 RP SIALIC ACID BINDING.  
 RX MEDLINE=95179521; PubMed=7533044;  
 RA Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,  
 RA Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;  
 RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new  
 RT family of sialic acid-dependent adhesion molecules of the  
 RT immunoglobulin superfamily";  
 RL Curr. Biol. 4:965-972 (1994).  
 RN [5]  
 RP BINDING TO SPN.  
 RX MEDLINE=21136329; PubMed=11238599;  
 RA van den Berg T.K., Nath D., Zillener H.J., Vestweber D., Fukuda M.,  
 RA van Die I., Crocker P.R.;  
 RT "CD43 functions as a T cell counterreceptor for the macrophage  
 RT adhesion receptor sialoadhesin (Siglec-1)";  
 RL J. Immunol. 166:3637-3640 (2001).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.  
 RX MEDLINE=96325385; PubMed=9660955;  
 RA May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;  
 RT "Crystal structure of the N-terminal domain of sialoadhesin in  
 RT complex with 3' sialyllactose at 1.85 A resolution";  
 RL Mol. Cell 1:719-728 (1998).  
 RN [7]  
 RP STRUCTURE OF 20-138 BY NMR, AND MUTAGENESIS OF TRP-21 AND ARG-116.  
 RX MEDLINE=99321481; PubMed=10393093;  
 RA Crocker P.R., Vinson M., Kelm S., Drickamer K.;  
 RT "Molecular analysis of sialoside binding to sialoadhesin by NMR and  
 RT site-directed mutagenesis";  
 RL Biochem. J. 341:355-361 (1999).  
 CC -1- FUNCTION: Macrophage-restricted adhesion molecule that mediates  
 CC sialic-acid dependent binding to lymphocytes, including  
 CC granulocytes, monocytes, natural killer cells, B-cells and CD8 T-  
 CC cells (By similarity). Preferentially binds to alpha2,3-linked  
 CC sialic acid. Binds to SPN/CD43 on T-cells. May play a role in  
 CC hemopoiesis.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and  
 CC soluble (isoforms 2 and 3).  
 CC -1- ALTERNATIVE PRODUCTS: Named isoforms=3;  
 CC Event=Alternative splicing; Name=1;  
 CC Name=1;  
 CC IsoId=Q62230-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q62230-2; Sequence=VSP\_002573, VSP\_002574;

CC Name=3;  
 CC IsoId=Q62230-3; Sequence=VSP\_002575, VSP\_002576;  
 CC -1- TISSUE SPECIFICITY: Expressed by macrophages in various tissues.  
 CC Highest expression in spleen and lymph node with lower amounts in  
 CC lung, liver, bone marrow, heart and skin. No expression in thymus,  
 CC kidney, brain or small intestine.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC  
 CC (sialic acid binding Ig-like lectin) family.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -1- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z36293; CAA85290.1; -;  
 DR EMBL: Z36233; CAA85268.1; -;  
 DR EMBL: Z36234; CAA85269.1; -;  
 DR EMBL: U92842; AAB95641.1; -;  
 DR EMBL: U92833; AAB95641.1; JOINED.  
 DR EMBL: U92834; AAB95641.1; JOINED.  
 DR EMBL: U92835; AAB95641.1; JOINED.  
 DR EMBL: U92837; AAB95641.1; JOINED.  
 DR EMBL: U92838; AAB95641.1; JOINED.  
 DR EMBL: U92839; AAB95641.1; JOINED.  
 DR EMBL: U92840; AAB95641.1; JOINED.  
 DR EMBL: U92841; AAB95641.1; JOINED.  
 DR PIR: S50065; S50065.  
 DR PDB: 1QFO; 16-APR-99.  
 DR PDB: 1QFP; 16-APR-99.  
 DR MGD: MGI:99668; Sn.  
 DR GO: GO:0016021; C: integral to membrane; ISS.  
 DR GO: GO:0005529; F: sugar binding; ISS.  
 DR GO: GO:0016337; F: cell-cell adhesion; ISS.  
 DR GO: GO:0007160; P: cell-matrix adhesion; ISS.  
 DR GO: GO:0006954; P: inflammatory response; ISS.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR Pfam: PF00047; Ig; 16.  
 DR SMART: SM00408; IGc2; 5.  
 DR PROSITE: PS00835; IG-LIKE; 14.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;  
 KW Immunoglobulin domain; Repeat; Alternative splicing; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1694  
 FT DOMAIN 20 1638  
 FT TRANSMEM 1639 1659  
 FT DOMAIN 1660 1694  
 FT DOMAIN 20 136  
 FT DOMAIN 153 235  
 FT DOMAIN 239 321  
 FT DOMAIN 326 406  
 FT DOMAIN 416 508  
 FT DOMAIN 509 594  
 FT DOMAIN 602 701  
 FT DOMAIN 704 781  
 FT DOMAIN 795 890  
 FT DOMAIN 894 972  
 FT DOMAIN 979 1078  
 FT DOMAIN 1080 1160  
 FT DOMAIN 1171 1263  
 FT DOMAIN 1244 1336  
 FT DOMAIN 1341 1438  
 FT DOMAIN 1441 1519  
 FT DOMAIN 1533 1626  
 FT SITE 827 829  
 FT DISULFID 36 166  
 BR SIALOADHESIN.  
 BR EXTRACELLULAR (POTENTIAL).  
 BR POTENTIAL.  
 BR CYTOPLASMIC (POTENTIAL).  
 BR IG-LIKE V-TYPE.  
 BR IG-LIKE C2-TYPE 1.  
 BR IG-LIKE C2-TYPE 2.  
 BR IG-LIKE C2-TYPE 3.  
 BR IG-LIKE C2-TYPE 4.  
 BR IG-LIKE C2-TYPE 5.  
 BR IG-LIKE C2-TYPE 6.  
 BR IG-LIKE C2-TYPE 7.  
 BR IG-LIKE C2-TYPE 8.  
 BR IG-LIKE C2-TYPE 9.  
 BR IG-LIKE C2-TYPE 10.  
 BR IG-LIKE C2-TYPE 11.  
 BR IG-LIKE C2-TYPE 12.  
 BR IG-LIKE C2-TYPE 13.  
 BR IG-LIKE C2-TYPE 14.  
 BR IG-LIKE C2-TYPE 15.  
 BR IG-LIKE C2-TYPE 16.  
 BR CELL ATTACHMENT SITE (POTENTIAL).  
 BR BY SIMILARITY.

```

FT DISULFID 41 98 BY SIMILARITY.
FT DISULFID 160 218 BY SIMILARITY.
FT DISULFID 263 306 BY SIMILARITY.
FT DISULFID 347 391 BY SIMILARITY.
FT DISULFID 434 492 BY SIMILARITY.
FT DISULFID 532 576 BY SIMILARITY.
FT DISULFID 625 685 BY SIMILARITY.
FT DISULFID 725 770 BY SIMILARITY.
FT DISULFID 813 872 BY SIMILARITY.
FT DISULFID 911 955 BY SIMILARITY.
FT DISULFID 1000 1062 BY SIMILARITY.
FT DISULFID 1102 1144 BY SIMILARITY.
FT DISULFID 1188 1236 BY SIMILARITY.
FT DISULFID 1276 1319 BY SIMILARITY.
FT DISULFID 1362 1421 BY SIMILARITY.
FT DISULFID 1462 1508 BY SIMILARITY.
FT DISULFID 1551 1610 BY SIMILARITY.
FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1459 1459 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1473 1473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 326 340 MAEYKMPAGEVLN -> SESWRLKGPVSGKH (in isoform 2).
FT VARSPLIC 341 1694 /Frida-VSP 002573.
FT VARSPLIC 1528 1598 Missing (in isoform 2).
FT VARSPLIC 1599 1694 /Frida-VSP 002574.
FT MUTAGEN 21 21 W->: LOSS OF SIMILIC ACID BINDING.
FT MUTAGEN 116 116 R->A: LOSS OF SIMILIC ACID BINDING.
FT MUTAGEN 116 116 R->L: 10-FOLD LOSS IN AFFINITY TO SIMILIC
Query Match 7.3% Score 96.5; DB 1; Length 1694;
Best Local Similarity 24.8%; Pred. No. 20;
Matches 75; Conservative 37; Mismatches 106; Indels 85; Gaps 17;

```

```

Db 284 LGV 286
LU HUMAN
ID LU HUMAN STANDARD; PRT; 628 AA.
AC P50895;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lutheran blood group glycoprotein precursor (B-CAM cell surface glycoprotein) (Auberger B antigen) (Fg/G253 antigen).
GN LU OR BCAM OR MSX19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
RC TISSUE=Placenta;
RX MEDLINE=95296337; PubMed=7777537;
RA Parsons S.F., Mallinson G., Holmes C.H., Houlihan J.M., Simpson K.L., Mawby W.J., Spurr N.K., Warne D., Barclay A.N., Anstee D.J.;
RT "The Lutheran blood group glycoprotein, another member of the immunoglobulin superfamily, is widely expressed in human tissues and is developmentally regulated in human liver."
RL Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
RN [2]
RP SEQUENCE OF 1-588 FROM N.A.
RX MEDLINE=95042297; PubMed=7954395;
RA Campbell I.G., Foulkes W.D., Senger G., Trowdale J., Garin-Chesa P., Reutig W.J.;
RT "Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers: a novel member of the immunoglobulin superfamily."
RL Cancer Res. 54:5761-5765(1994).
RN [3]
RP CARBOHYDRATE-LINKAGE SITE ASN-439.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Abersold R.;
RT "Identification and quantification of N-linked glycoproteins using hydrazone chemistry, stable isotope labeling and mass spectrometry." Nat. Biotechnol. 21:660-666(2003).
CC - FUNCTION: PROBABLE RECEPTOR. MAY MEDIATE INTRACELLULAR SIGNALING.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION (HIGHEST IN THE PANCREAS AND VERY LOW IN BRAIN). CLOSELY ASSOCIATED WITH THE BASAL LAYER OF CELLS IN EPITHELIA AND THE ENDOTHELIUM OF BLOOD VESSEL WALLS.
CC - DEVELOPMENTAL STAGE: Is under developmental control in liver and may also be regulated during differentiation in other tissues.
CC - UPREGULATED following malignant transformation in other cell types.
CC - POLYMORPHISM: LU is responsible for the Lutheran blood group system.
CC - SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC - SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
CC - DATABASE: NAME=Blood group antigen mutation database; NOTE=Lutheran (Lu) blood group system; WWW="http://www.bdoc.aecom.yu.edu/bgmutc/lutheran.htm".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC EMBL; X83425; CAA58449.1; -
CC EMBL; X80026; CAA56327.1; -
CC PIR; I37202; I37202.
CC PIR; I38000; I38000.

```